

Paying the price of health care

The British government's plans to reorganize its public health service are not as fearsome as its critics say, but endanger medical research — and the civility that health care has brought to Britain.

ABOUT the only certainty in the business of health care is that the cost will always increase more quickly than it can be afforded. The explanation is both simple and familiar. Broadly speaking, better health implies longevity, meaning that a person's chance of living to debilitated old age is increased, so that the proportion of people needing care of some kind grows with the passage of time. So do people's expectations. One decade, they may be reconciled to the view that the best treatment for heart disease is its avoidance (exercise, sensible diet and so on), the next they will be looking for coronary by-pass operations when they need them, which means that age-specific unit costs also tend to increase. Of course, there are some countervailing tendencies. As healthier people live longer, morbidity and thus the need for health care at a given age may decline. In most industrialized countries, this has been one of the outstanding demographic phenomena of the past few decades.

And the unit cost of even coronary by-pass operations declines as more of them are performed, although that does not imply that the total cost of heart surgery stops growing. But experience seems to show that tendencies towards increasing costs predominate.

These familiar considerations are essentially why the British government has spent more than a year deciding how best to put its National Health Service on a different and, with luck, more stable financial footing. That, at least, is the government's view of its intentions. Its critics would prefer that it were understood that the year has been spent searching for ways of reducing the cost to the public purse of the National Health Service (NHS), now twice the annual defence budget. The truth is that the issue is not simply economic, but political. In general terms, the British government believes that public spending should be a minimum; the opposition parties that the present government spends too little on helping society tick over. The difference of opinion also goes back to the origins of the NHS, which was founded (in 1947) by a Labour government and which has been managed for more than half of the intervening time by governments of a different persuasion. But the outcome of the government's study, a set of recommendations for discussion, will disappoint its friends as well as its opponents.

Inspiration

The British NHS has a special place in recent social history. Its inspiration, in the impoverished 1920s and 1930s, was the belief that people should not be denied access to medical care for lack of funds. The method chosen in 1947 to put this principle into practice was to declare that medical treatment should be free of charge to all who asked for it. Aneurin Bevan, the Cabinet minister who pushed the legislation through the British Parliament, believed there would be such a rapid improvement of the general health of the British population that the costs of free treatment would at worst remain static. Forty years of experience have shown that the NHS has had a powerful influence on the temper of British society. The elderly and the chronically sick have been cared for without regard for their personal resources. As it happens, other sections of British society have delighted in the NHS: a little to the present British

government's embarrassment, many of its own supporters are also among the strongest supporters of what has been called elsewhere a system of "socialized medicine".

The upshot of the government's new proposals for the NHS is not radical but managerial. The government has backed away from its earlier hopes that it might find ways of shedding part of the cost of the NHS to other independent sources of funds — the pockets of patients through compulsory health insurance, for example — and has opted instead for a strategy for making the NHS much more efficient. Major hospitals are to be allowed to contract out of direct control by regional health authorities, while family physicians in sufficiently large practices will be encouraged to work within an annual budget and allowed to keep half of whatever costs they save. The opposition parties regard this as a first step towards complete devolution of government responsibility for the NHS, but that is an issue for a future Parliament — and the proposed reforms will not be complete until 1992 or later. The more serious difficulty is that the changes, if they work, will yield a once-and-for-all benefit, but not solve the long-term problem occasioned by the disproportionately rapid increase of costs.

Danger

Meanwhile, the changes now planned are likely to damage some important functions of the NHS as it has evolved, responsibility for research and innovation in particular. Under the present system, the Medical Research Council (MRC) has funds for the support of basic medical research, but parts of NHS, in partnership with the MRC or otherwise, supports the development of new techniques by meeting the costs out of current budgets. What will happen when the major hospitals in which this work inevitably concentrates will be autonomous but driven by the need to balance their books, and to keep their costs within reasonable limits so as to compete with other hospitals? There is at least a danger that research projects will go to the wall. Before the changes planned have made much headway, the government may find itself having to create a second source of research funds — or contemplate the possibility that British medicine will lag behind that elsewhere?

The other obvious danger is that the new regime will prove one in which the people most in need of treatment will find themselves, for lack of personal resources, at the end of whatever queues there are. Although the government says that there is no intention of departing from the principles on which the NHS is run at present, last week's proposals say nothing about the mechanism by which this is to be done. The elderly and the chronically sick will be, in relation to the new NHS, in much the position of the consumers of the newly privatized nationalized industries — less able to exercise their rights as consumers than they should be. That is why even the modest changes in the management of the NHS now proposed demand ways in which the interests of the patients least able to look after themselves can be safeguarded.

If such a mechanism is not provided, the government will find that — contrary to its expressed intentions — the civility that the NHS has brought to British life will be put in hazard. □



NIH report not the final word on Baltimore case

- Call for "letter of correction" to *Cell*
- Another congressional hearing to come?

Washington

THE US National Institutes of Health (NIH) have at last released the final report of their investigation of possible scientific misconduct in the publication of a paper in *Cell* (45, 247-259; 1986). The report absolves the authors of the paper of misconduct, but the dust has not yet settled.

The report is almost identical to a draft circulated last November which concluded that there is no evidence of misconduct, but that the published paper contained "significant errors of misstatement and omission", as well as "lapses in scientific and inter-laboratory communication". The principal authors are Thereza Imanishi-Kari and Nobel laureate David Baltimore, together with four others (see *Nature* 336, 505; 1988).

The handling of the investigation has become a contentious issue between a House of Representatives congressional committee and NIH. Before the publication of the report, NIH asked the subcommittee on oversight and investigations whether it had further evidence bearing on the inquiry. In a letter dated the day before the report was released, the chairman, Representative John Dingell (Democrat, Michigan), said it would be "improper" for the committee to advise NIH, but that its function was to monitor NIH's conduct of inquiries such as this.

Complaints about the *Cell* paper, which reports unexpected immunological characteristics of transgenic mice, were first made in 1986 by Margot O'Toole, a postdoctoral researcher in Imanishi-Kari's laboratory at the Massachusetts Institute of Technology, and became public when Walter Stewart and Ned Feder, two NIH scientists known for their investigations of scientific misconduct, took up the case. After two internal inquiries (one at Tufts University) and two years of debate, NIH established a panel of three to investigate the affair and to report.

But by that stage (early last year) the dispute had been taken up by a powerful congressional subcommittee headed by John Dingell, who "borrowed" Stewart from NIH to investigate charges that NIH are soft on scientific misconduct. And in April Dingell's subcommittee held a congressional hearing on "Fraud In NIH Grant Programs" in which the controversy

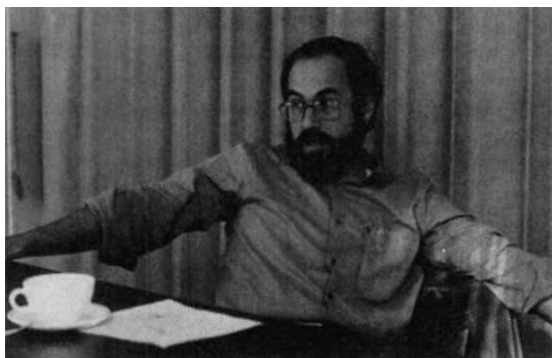
over the *Cell* paper featured prominently.

The panel's report concludes that there are errors sufficient to warrant a letter of correction in *Cell* in addition to a correction already published by the authors of the paper in November. NIH director James Wyngaarden, in a letter to Imanishi-Kari and her co-authors, asks that they send him a copy of the correction before submission so that NIH may consider its "completeness and accuracy". The report also recommends submission to *Cell* of a brief report on the "problems" associated with the relative sensitivity and specificity of the reagents and assays used in the experiments.

Reaction to the report is varied. Baltimore says he feels "vindicated" because the report puts to rest all accusations of improper conduct and supports the central conclusions of the original paper.

But, while accepting the panel's recommendation to provide *Cell* with a discussion of the reagents and assays (subject to the proviso that the journal agrees to publish it), Baltimore and two of his co-authors, in a letter to NIH responding to the draft report, say that there are no "errors" that need correction, only "differences of judgement".

Although Baltimore says that their rebuttal "seems to have passed NIH by without notice", Katherine Bick, NIH deputy director for extramural research, says Baltimore's concerns were carefully considered, but that they did not change the investigating panel's opinion that a



David Baltimore: no errors, only "differences of judgement".

correction is needed.

O'Toole, who submitted detailed criticisms of the draft report, still believes that NIH's investigation was "wholly inadequate". None of her criticisms prompted any substantial change in the final report, one of which is that the panel failed to interview Stewart and Feder, who spent

Fifty years of fission



THE now-familiar mushroom cloud — first seen in 1945, six years after fission was first observed. To mark the fiftieth anniversary of the discovery of the fission of uranium, *Nature* will be publishing a series of articles on fission and its consequences. The first appears on page 499.

considerable time and effort probing into the affair. O'Toole says she was asked to clarify some of Stewart and Feder's criticisms for the investigating panel.

NIH did seek the opinion of the Dingell subcommittee, to which Stewart has been seconded from NIH. In a letter dated 20 December, William Raub, NIH deputy director, asked Dingell if the subcommittee had "any information ... inconsistent with the confidential draft report of the NIH panel of scientists".

In a reply dated 31 January (the day NIH released the report), Dingell pointed out that, as the subcommittee has responsibility for oversight of NIH activities, it would be "improper" to provide advice and help in the completion of the NIH investigation. Dingell said he regards NIH's performance in this case as a "crucial test of their ability to deal with cases of questioned science". Dingell pointed to some of the criticisms of the draft report by Baltimore and O'Toole and said he "trusts that these matters will be resolved" in a "factually unimpeachable manner" in the final report. Peter Stockton, a member of the Dingell committee staff, says the final report is "adequate in as far as it goes". The subcommittee is particularly pleased that Wyngaarden is insisting on written corrections for the errors in the *Cell* paper, and chastises the authors for not responding sooner to initial criticism of their paper. But Stockton says this is "not the end of the Baltimore issue" and that the subcommittee will "probably" hold another congressional hearing.

David Swinbanks

West German government reeling from export disclosures

Munich

THE mushrooming story of West German participation in the alleged chemical weapons factory at Rabta in Libya and in biological and atomic weapons programmes in Iraq and India has plunged Christian Democrat Helmut Kohl into the gravest crisis of his six-year chancellorship.

On 30 January, the news magazine *Spiegel* revealed that the West German subsidiary of the US-based Sigma chemical company had delivered 100 mg of the deadly mycotoxins HT-2 and T-2 to Iraq in 1987 for DM 60,000. Government spokesman Friedhelm Ost defended the decision not to require an export licence, saying that the amount was so small that the toxins could only have been used for scientific purposes. A toxicologist at the Walter-Straub-Institute in Munich said that 100 mg was indeed too little toxin to use for biological weapons, but added that research on antidotes to the toxins or on other defensive measures would have been possible.

On the same day, *Spiegel* revealed that Degussa AG of Frankfurt had exported 95 kg of beryllium to India in 1984. Beryllium can enhance the power of atomic bombs, although it also has non-military uses. Much of the beryllium sent to India had been imported from the United States, and US officials said that West Germany should have applied for a licence to re-export it. The government says it approved of the export because it had assurances that the beryllium would be used for non-nuclear purposes. The beryllium was reportedly delivered to the Bhabha Atomic Research Center in Trombay.

Potentially even more damaging to the government was the disclosure that state-owned Salzgitter Industriebau AG had apparently known since 1985 that

Imhausen's "pharmaceuticals factory", for which it was creating blueprints, was destined for Libya and not for Hong Kong. This emerged from documents discovered by Salzgitter officials and turned over to prosecutors on 29 January.

The revelation was the first potential "evidence that would stand up in court" linking privately owned Imhausen-Chemie GmbH with the Libyan plant. On 10 January, Chancellor Kohl had vigorously denied that such evidence existed. Officials at Salzgitter say that they were "fooled" by Imhausen and that they did not know the real destination of the "pharmaceuticals factory".

The government has intensified the search for evidence in the Rabta case and proposed harsher sanctions for West German citizens abroad who are found to be involved in the construction of biological or chemical weapons factories.

There has yet to be any serious discussion about changing the West German export laws. The current system allows all exports except those specifically restricted; the US system forbids all exports except those specifically allowed.

But even if new laws are enacted quickly, the government's credibility may have sustained irrevocable damage. The durable coalition between Christian Democrats (CDU) and Free Democrats is now fragile. Other developments, especially a devastating defeat for the CDU in the West Berlin state elections on 29 January, have contributed to the malaise.

Kohl dispatched two top envoys to Washington last week to mend fences and curry favour with the Bush administration. He must now prepare answers to 49 questions asked by the opposition regarding government behaviour in the Libya affair.

Steven Dickman

Cover story hits the wrong target

Washington

THE cover of a scientific products catalogue from PGC Scientifics Corporation of Gaithersburg, Maryland, has elicited howls of protest from researchers, particularly at NIH who are appalled by its inappropriateness. The cover depicts a man and a woman dressed in white laboratory coats — but made up to look like film versions of "mad scientists" — attempting to force a blue liquid down a white rat's throat.

Robert Whitney, head of the NIH division of research services, says his office has received at least a dozen calls from NIH researchers complaining about the cover. He says the cover conveys "absolutely the wrong message" about researchers and how they work, especially with regard to animals. "Whoever came up with the advertisement must have been living in Siberia for the last 10 years. It's absolutely inexcusable", says Whitney.

NIH is planning to write a letter of protest to the company, but is not contemplating any boycott or sanctions.

Gary Bechanan, the marketing manager of the company, says the cover was intended as a humorous portrayal of the common stereotype of scientists, and that it was "not our intent to imply animal abuse". He and the president of the company are responding personally to all complaints about the cover, and an apology letter will be printed in the new edition of the catalogue, to be released in August. In the meantime, the current catalogue is being reprinted with an old cover which depicts a cityscape made from labware.

Joseph Palca & Carol Ezzell

The cover you won't see . . .



US science literacy could be better

Boston

STUDENTS from the United States and French-speaking Ontario in Canada have been ranked last in a survey of mathematics and science performance of 13-year-olds from six industrialized nations, according to a report released last week by the Educational Testing Service, based in Princeton, New Jersey. The report's findings, called a "national tragedy" by the new US Secretary of Education, Lauro Cavazos, will further fuel growing concern about mathematics and science literacy in the United States.

The survey tested 24,000 students from Britain, Ireland, Spain and South Korea, as well as the United States and four Can-

adian provinces. South Korean students scored highest in both mathematics and science, followed closely by students from British Columbia, Canada. According to the Educational Testing Service, the survey is the first large-scale international comparison of students in this age group.

Interestingly, the findings tend to refute previous studies suggesting that there is a gender gap between boys and girls in mathematics performance. Only in Korea and Spain did boys significantly outperform girls in this test. In the science portion of the test, however, boys did better than girls everywhere except in the United States and the United Kingdom.

Seth Shulman

Ozone hole looms large

Washington & London

BAD news is good news for four research teams conducting investigations into the depletion of stratospheric ozone in the Arctic. Polar stratospheric clouds (PSCs), containing the ice crystals thought to cause formation of active chlorine that destroys ozone, are forming in large numbers in the Arctic, which could result in a sizeable depletion of Arctic ozone this year. But they are forming within range of measuring equipment and aircraft, and data collected from these expeditions should confirm whether atmospheric ozone depletion is occurring in the Arctic.

The Canadian station at Alert at 82.6° north latitude was able to record the first data. According to Wayne Evans of the Canadian Atmosphere Environment Service, the polar vortex was centred over Alert for much of January. The Canadian team has been using balloon-launched ozone sondes, backscatter sondes and water-vapour sondes to measure ozone levels and cloud-ice concentrations.

The location of the vortex was frustrating a team assembled by the US National Aeronautics and Space Administration working out of Stavanger, Norway (see *Nature* 337, 3; 1989) because it was out of range of the high-altitude ER-2 aircraft used by the team to fly into the stratospheric clouds. The US-led team was also experiencing difficulty with accurate weather forecasting for flying into the North Sea, according to NASA spokesman Charles Redmond. But now, according to Evans, the vortex has shifted towards Franz Josef Land on the Soviet side of the Arctic, and Redmond says both the ER-2 and the lower-flying DC-8 have completed 11 missions into areas where there are PSCs. Another four flights are planned before the team leaves Norway.

The movement of the vortex has put Soviet researchers in a prime position for studies from their base on Heiss Island, one of a group of islands in Franz Josef Land. According to Redmond, they are carrying out Dobson measurements of ozone and are believed to be sending balloons and aircraft into the vortex. Direct read-outs of their ozone soundings are being received at Stavanger.

Another international team, including groups from West Germany, France and the United States base at Kiruna in Northern Sweden, has also been busy making measurements in the polar vortex. They have made a number of successful balloon and rocket flights into the PSC-populated area, allowing measurements of ozone, stratospheric cloud particles and nitric acid.

Joseph Palca & Philippa Lloyd

Political tide turns as threats to global climate develop

Washington

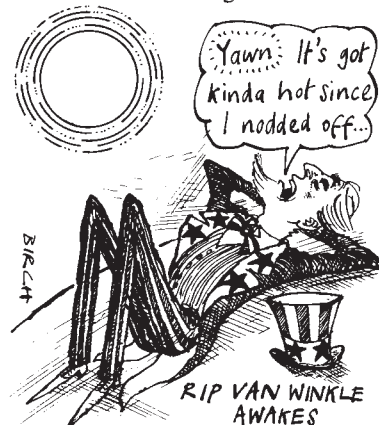
THE new US Secretary of State, James Baker, last week declared that the time has come for political action on the problem of global climate change. Baker's statement, made at the first meeting of a working group of the Intergovernmental Panel on Climate Change, is being interpreted as a major shift in US foreign policy. But Baker's call for action seems to have impelled the working group chiefly to follow US proposals and to establish bureaucratic procedures for future work.

Many of the themes at the meeting are familiar: the need for more research on renewable energy sources, renewed attention to nuclear power generation to reduce greenhouse-gas production, calls for added restrictions on ozone-damaging chlorofluorocarbon emissions and a campaign to halt deforestation.

But the draft report produced at the end of the closed three-day meeting does not address these matters, and instead outlines responsibilities for a steering group and four subgroups established at the meeting (energy and industry; other human activities; coastal zone management; and resource use and management). The only 'action' taken was a decision to ask the United States and the Netherlands to develop preliminary emission profiles for gases such as carbon dioxide and chlorofluorocarbons between now and the year 2100 for three scenarios: carbon dioxide equivalent doubling by 2030; doubling by 2050; and doubling by 2070. Some delegates complained that the US proposal seemed at odds with Baker's statement

and that it followed the former Reagan administration's practice of advocating more research and data collection before taking action. But Frederick Bernthal, assistant secretary of state, Bureau of International Environmental and Scientific Affairs, says the US proposal was "more ambitious" than the final draft adopted.

Meanwhile US politicians are moving ahead with their own agenda. Last week,



Senator Timothy Wirth (Democrat, Colorado) reintroduced legislation to establish a national energy policy to reduce emissions of greenhouse gases.

Late last month, Senator Ernest (Fritz) Hollings (Democrat, South Carolina), introduced the National Global Change Research Act to coordinate federal research on greenhouse warming, ozone depletion and other long-term environmental changes.

David Swinbanks

Mounting opposition to UK student loans

London

THE British government's plans to supplement student maintenance grants with loans have run into more opposition last week from the Committee of Vice-Chancellors and Principals (CVCP), which represents the common interests of British universities.

In a formal criticism of the loans scheme, announced only last November, the committee calls the plan "inadequate and unjust", fearing that students will be frightened away from longer courses (such as specialized engineering and teacher training) and from those leading to poorly paid vocations. It says the proposals are inconsistent with the need for the wider access to university education to which the government has recently been converted.

The committee says that it would in any case prefer a scheme in which graduates from higher education are taxed for a fixed

period of years (provided that their salaries exceed the average), on the grounds that contributions by ex-students should be related to the value of their qualifications and not to the cost of providing them.

Commercial banks, which the government was relying on to administer the loans, also oppose the plan. The Department of Education and Science assumed that banks would subsidize the scheme in their enthusiasm to gain student customers. But the chief economic adviser to Lloyds Bank, Christopher Johnson, said this week that banks might actually lose student customers by becoming "debt-collectors" for the government. He proposes instead that conventional commercial loans be made to parents, with tax-free interest, and that both the government grant and parental contribution be increased by 18 per cent.

Christine McGourty

Children infect mothers in AIDS outbreak at a Soviet hospital

Moscow

ALMOST the whole Soviet Union has been shocked by the revelation that a single hospital at Elista, the capital city of the Kalmyk autonomous republic (population 344,000), has managed to infect no fewer than 27 young children and five of their mothers with the human immunodeficiency virus (HIV).

At the end of last year, there were only about 120 officially registered AIDS cases in the Soviet Union, of whom only six had clinical symptoms of disease.

The immediate cause of the Elista outbreak is the criminal negligence and irresponsible attitude of the medical staff at the children's hospital at Elista, where syringe needles were reused without adequate sterilization.

The outbreak was first recognized late last November, when a new-born baby at the hospital tested positive for HIV. Following established procedures, the case was reported to the Central Institute for Epidemiology in Moscow, whose specialized AIDS laboratory collects information and also examines all suspected cases of HIV infection. Tests on the baby from Elista confirmed suspicions, but both parents tested negative for HIV.

Almost at the same time, the blood donor service at Elista found another virus carrier by HIV screening (routine since the beginning of 1988) — a young woman with a recently dead baby. Although her contacts revealed no further infection, it emerged that her dead child had spent almost a month at the Elista hospital, which prompted an urgent investigation by the Moscow AIDS laboratory.

The outcome stunned the investigators, who examined serum samples from all the children treated at the hospital at the same time as the dead child and the newly infected children. No fewer than 4 out of 19 tests were positive for HIV.

Vadim Pokrovsky, head of the AIDS laboratory, promptly flew to Elista with a team to collaborate with the republic's public health services in screening all children who could have had contact with those infected. One complication is that most child patients at the Elista hospital suffer from chronic diseases and are referred there only when treatment at other general hospitals has been ineffective. But 27 infected children have now been identified.

Knowing that HIV infection can be transferred from mother to new-born child, the investigators expected to find as many infected women as children. But only four of the 27 mothers were found to be infected. So how could the children of uninfected mothers acquire HIV? Most

probably, by the use of non-sterile medical instruments and serious breaches of the rules of hygiene at the hospital.

Meanwhile, the medical commission has concluded that the five infected mothers (the four now found and the mother of the dead child) must have been infected by their children. The 'external' contacts of the four women are acknowledged to be clear, and screening of 12,000 Elista residents has turned up no further cases. One possibility is suggested by the observation that many of the children treated at Elista are weak, with stomatitis and bleeding gums, and that HIV may be transmitted from child to mother in breast-feeding via cracks in the nipple.

While it is too soon to draw final conclusions about the Kalmyk AIDS outbreak, some specialists expect that two or three times as many AIDS cases will eventually be found as are now known.

But the Kalmyk situation says much about the drive against AIDS in the Soviet Union. It is plain that the headstart of several years — the time-lag between the

appearance of infection in the United States and the Soviet Union — has not been used to mount effective preparations against an epidemic. While traditional stereotypes have been a brake on the spread of infection, the attitudes of the health authorities have also been an obstacle.

With the restructuring of the health services now under way, top Soviet officials are now in close touch with the World Health Organization and are seeking to erect reliable barriers against an AIDS epidemic. A uniform state-wide programme has been adopted, screening is under way, laboratories are being set up and Soviet-made instruments and diagnostic reagents are being developed. But the task is complicated by laggard industry, the sometimes low skill of medical personnel and the often inadequate state of public health information — the word "condom" made its appearance in the press barely six months ago.

Direct responsibility for the AIDS outbreak at Elista will be determined by the courts, but it must be hoped that the legal investigation will be a turning point for Soviet attitudes towards the AIDS problem

Viktor Belitsky
Novosti

Unproved AIDS drug is made available for limited use

Boston

THE Food and Drug Administration (FDA) has this week granted limited approval for a drug believed to be effective in preventing *Pneumocystis carinii* pneumonia, the leading cause of death in AIDS patients. The decision makes good the FDA's commitment last year to bring promising drugs more swiftly to patients with AIDS or other life-threatening illnesses.

The drug, an aerosol version of pentamidine, has been shown in a preliminary and as yet unpublished clinical study at San Francisco General Hospital to inhibit the recurrence of *Pneumocystis* pneumonia. FDA's authorization allows the use of the drug in "secondary prophylaxis" in patients who have already suffered a bout of the pneumonia. The FDA is also authorizing its use in patients who have not yet contracted the pneumonia provided that their T4 blood-cell count is less than 200 per millilitre of blood. This decision is based solely on preliminary data from a federal study showing that patients with low T blood-cell counts face a particularly great risk of contracting the pneumonia.

Jerome Groopman, an AIDS researcher at the New England Deaconess Hospital in Boston, says that FDA's decision on primary prophylaxis is "important and

correct" and that it shows the FDA's willingness to act on the basis of clinical data "which makes sense in practice without necessarily conforming to the highly stringent requirements" of final approval.

The procedure followed in approving the aerosol form of pentamidine, known as the treatment investigational new drug (IND) procedure (see *Nature* 326, 536; 1987), has been used previously to offer drugs to patients before final licensing. Azidothymidine (AZT), the only drug licensed for use against the AIDS virus HIV, was also at first released as a treatment IND. But this is the first time that the manufacturer is likely to take advantage of a provision allowing it to recover the costs of the drug's dissemination during this period. An announcement about the financial and logistical details, including a special telephone number which physicians and hospitals can call to order the drug, is expected shortly by LyphoMed Inc. of Rosemont, Illinois.

The aerosol version of pentamidine has hitherto been available only in research programmes, most of which are centred on San Francisco and New York. But pentamidine in intravenous form, or by subcutaneous injection, is already licensed and marketed widely as a treatment for *Pneumocystis* pneumonia.

Seth Shulman

More insults fly but Japan's 'scientific' whaling goes on

Tokyo

THE week-long Antarctic duel between the Japanese whaling fleet and the Greenpeace ship *Gondwana* has left Japanese government officials furious. "Terrorism on the high seas" is how one Ministry of Agriculture, Forestry and Fisheries official angrily labelled Greenpeace's actions after the ice-strengthened *Gondwana* collided with the Japanese factory ship *Nisshin Maru No 3* last Tuesday.

Japan's Foreign Ministry has taken a less abrasive line, saying that "terrorism is not yet the legitimate wording" for Greenpeace's attempts to halt the killing of minke whales by sending in rubber dinghies between the catcher boats and the whales. The Foreign Ministry has asked for help from the Netherlands where the *Gondwana* is registered.

The Foreign Ministry contends that, as the Japanese "research" fleet's activities are "in line with International Whaling Commission (IWC) stipulations", disruptive activities are "unjustifiable". But Greenpeace argues that by labelling its whaling activities as "research" Japan is exploiting a loophole in IWC rules.

Under pressure from the United States, Japan agreed to give up whaling from

1988, in line with an IWC moratorium on commercial whaling. The moratorium was adopted because the IWC considers that the survival of some species is in doubt.

But a year before it was due to quit, Japan came up with a proposal to begin 'scientific' research whaling. Under Article 8 of the International Whaling Convention, member nations can kill as many whales as they like for scientific research purposes. Research proposals must be submitted to the IWC's scientific advisory committee but the committee's advice can be ignored.

Kazuo Yamamura, a spokesman for the Institute of Cetacean Research, responsible for managing Japan's 'scientific' whaling under government contract, argues that Japan's whaling is true scientific research. All commercial whaling companies have now closed down, he says, and the research whaling, which cost around 1,700 million yen (\$1.4 million) in 1988, cannot cover its costs from sale of whale meat. Substantial government grants and donations are necessary.

Although critics agree that Yamamura may be technically correct, they argue that by keeping the whaling fleet in action and

restaurants stocked with whale meat, Japan's research whaling is just a way to maintain readiness for the resumption of full-scale commercial whaling operations. But Yamamura contends that Japan is obeying the commercial whaling moratorium and would begin full-scale whaling only if the IWC lifts the moratorium after the comprehensive assessment of whale stocks due for 1990.

So far, at least, the majority of members of the IWC scientific advisory committee do not agree. Japan's original plans for scientific whaling were criticized as



A Greenpeace protester makes a point — by trying to come between whaler and whale.

unscientific and unhelpful to the IWC goal of rationally managing whale stocks. In an unprecedented move, the IWC passed a resolution by postal ballot in February 1988, asking Japan to stop whaling until the scientific merits of the programme could be debated. Japan ignored the ballot and continued its 'research'. But it was punished by US action that continued Japan's zero fishing quota in US coastal waters.

Since then, the first year's data have been gathered and the fleet has reached the Antarctic to begin its second season of sampling. Some 40 minke whales had been killed by the time the *Gondwana*, en route to supply the Greenpeace base in Antarctica, came upon the fleet in the Ross Sea. Although the harassment slowed the catch, the Fisheries Agency says the target of 300 minke whales can still be met before April.

Data amassed so far prove the success of the sampling method, according to the Fisheries Agency. But the data have yet to be independently assessed by the IWC scientific committee.

Even if scientific criticism is forthcoming, it may not shake public support

Nuclear power loses popularity in Japan

Tokyo

THE number of anti-nuclear protest meetings has reached an all-time high in Japan, according to National Police Agency figures released on Saturday, and has more than doubled since 1987.

Until the Chernobyl accident, the Japanese nuclear power industry could rely on support from the public, with the majority of citizens backing the building of more nuclear plants (see *Nature* 318, 595; 1985). But now surveys show supporters of nuclear construction are outnumbered two to one, with opposition particularly strong in areas near proposed construction sites.

Atomic Energy Commission officials have hit back by stressing the safety of Japan's nuclear plants and the environmental pollution caused by coal- and oil-fired plants. And during the past year government and the power industry have poured extra funds into pro-nuclear publicity campaigns. But the increased strength of the opposition movement has meant bigger publicity for mishaps at nuclear plants. Last week, an announcement that a pump failure had closed the Fukushima No 2 nuclear plant quickly led to predictions that a core melt-down could have resulted if the failure had not been spotted in time.

Opposition groups are also growing more sophisticated. In Tottori prefecture, campaign groups are buying up small parcels of land at possible future plant sites to block construction. And in the northern island of Hokkaido, protesters succeeded in collecting the signature of more than two per cent of the local population at the end of last year, forcing the regional government to consider a plebiscite on the opening of a new nuclear plant.

Alun Anderson

for whaling in Japan, where the media portray the problem as largely cultural, with Japan as a misunderstood victim of Western values. Attempts to ban commercial whaling are described as "an undeniable insult to the Japanese culture and to the Japanese people" in *Mainichi Daily News*. In *Daily Yomiuri*, a Fisheries Agency official attributes "US citizens" with "special sentiments for whales"; they are equated with religious beliefs that should not be imposed on Japan.

The long-term future of Japan's offshore whaling, however, will largely depend on the IWC's comprehensive assessment of whale stocks. If the IWC rules that commercial whaling could endanger whale populations and extends the moratorium, Japan will have to decide whether to keep scientific whaling going, to abandon whaling altogether, or to pronounce the IWC as biased and leave the convention, as threatened in the past.

Alun Anderson

Wrecked ship causes damage to Antarctic ecosystem

Berkeley & Washington

THE extent of environmental damage to the Antarctic sea and coastline from the 28 January wreck of an Argentine supply ship is still unclear. The ship, *Bahia Paraiso*, was carrying an estimated 250,000 gallons of diesel fuel when it ran aground near a US National Science Foundation (NSF) research station on the Antarctic Peninsula, 600 miles south of Cape Horn.

Two miles from Palmer Station, where the 81 tourists on board had visited the US scientists, the ship struck an underwater rock which tore a 30-foot gash in its hull. Staff from Palmer, with help from a number of passing tourist ships, rescued the crew and passengers, but the ship was abandoned. Fuel gushed out forming a slick several centimetres deep, but after

years. Of immediate concern is an Adélie penguin rookery, one of the largest in the area and home to 10,000 breeding pairs. A rapidly conducted survey reported by DeLaca revealed that 38 out of 40 inspected penguins were coated in fuel, and that the mortality rate among skua chicks was perhaps as high as 60 per cent.

The spill was also badly timed for the krill breeding cycle. Young krill hatched in deep water will be rising to the surface in mid-February, so will be exposed to the toxic fuel. DeLaca said that krill in shallow waters near Palmer Station are visibly affected by floating oil, and that several thousand dead krill have been washed up on the beaches, where they are quickly eaten by the penguins.

Diesel fuel presents different problems from a crude oil spill, said David Kennedy, scientific support coordinator for the National Oceanic and Atmospheric Administration's Hazardous Materials Response Branch in Seattle. It does not persist as long as crude oil, but is more toxic. "Diesel, if it hits a biologically sensitive area, has a tendency to kill quickly", said Kennedy, "but not to stay and affect

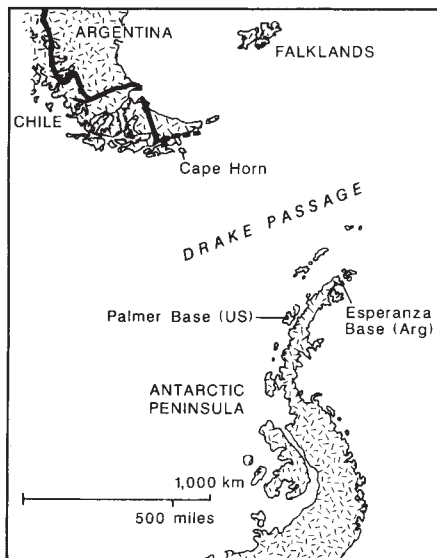
future generations". Cleanup, which involves containment of the spill, followed by soaking or skimming it off the water, will be aided by the relatively warm summer weather and lack of ice floes. Warmth and sunlight will help oxidize and evaporate the uncontaminated fuel.

Research activities at Palmer Station are likely to turn to a study of the effects of the spill on the ecosystem, information that may be valuable, should Antarctic ship traffic continue to increase.

Members of the Environmental Defense Fund (EDF), who happened to be on the US tour ship that participated in the rescue, have used the incident to underscore their concern about the environmental hazards of increased ship traffic and human activity in the Antarctic, and their opposition to a lifting of the current moratorium on mineral and petroleum exploitation.

DeLaca is concerned that the increasing numbers of tourist visits to the area rely, in such an emergency, on the limited logistical support that scientific bases can supply. Although Palmer Station staff coped successfully with the ship's passengers, and have been able to gather up some of the wreckage, they would have been helpless had the accident occurred a only few more miles away.

Marcia Barinaga & David Lindley



The oil spill near Palmer Station was the first major ecological incident in Antarctica.

this had subsided the ship broke loose of the rock and floated towards the shore, where it spilled more oil along the coast before finally coming to rest. Speaking from Palmer on 3 February, NSF chief scientist Ted DeLaca said that fuel oil was still leaking out at about two gallons a minute. A US Navy ship carrying oil containment booms and recovery equipment was expected to arrive late on 6 February, but in the meantime, an Argentinian vessel arrived last Saturday and picked up salvageable material. Divers from the ship reported that about one-third of the *Bahia Paraiso's* fuel oil was still in the tanks.

The day after the wreck, researchers at Palmer Station were finding krill and limpets that had been killed by the highly toxic fuel. Polly Penhale, NSF programme director for polar biology, said the spill has endangered the ecological research that has been conducted there for 20

New Zealand activists challenge US Navy astronomy station

Munich & Washington

A LONG-RUNNING dispute over a US Navy observatory in New Zealand is likely to flare up again when the New Zealand government considers peace activists' demands that the observatory be closed.

The US Naval Observatory built the Black Birch observatory on New Zealand's South Island to improve measurements of the fundamental astrometry of the southern sky. Black Birch replaced an older Naval Observatory station at Leoncito, Argentina. According to James Hughes, director of the astrometry department at the Naval Observatory in Washington, DC, the South Island location is better situated for meridian circle measurements, being nearly 10° further south than the Argentine facility.

But protests against the Black Birch observatory have "plagued" two New Zealand governments since construction began in 1982, according to an official in New Zealand's Ministry of Foreign Affairs. Peace activists claim that the observatory should be closed, and cite a clause in the 1987 New Zealand Nuclear-Free Zone Law which prohibits New Zealand citizens or residents from "aiding, abetting or procuring any person to manufacture, acquire, possess or have

control over any nuclear explosive device".

The leaders of the protest, Scientists Against Nuclear Arms and the Peace Movement Aotearoa, claim that data from the observatory will be used for military purposes. The protesters have called on the government to place the station under civilian management, allowing the current astrometry programme to continue, or, failing that, to close it down.

Hughes says the data being gathered at Black Birch will be used to improve star catalogues, and will be freely available to anyone. But protesters say the observatory provides data for the development of the next generation of Trident submarine-based missiles and for the Strategic Defense Initiative.

The protest movement gained momentum on 29 November, when a public advisory committee established under the Nuclear-Free Zone Law demanded that the New Zealand government reconsider its approval of the project.

A New Zealand Foreign Ministry official said it would be a "surprise" if the government did not respond to the request that Black Birch be shut down soon, possibly at the next advisory committee meeting on 21 February.

Steven Dickman & Joseph Palca

Europe launches new-technology education project by satellite

Paris

EUROPE last week joined the world of higher education by satellite television with the launch of the European Programme of Advanced Continuing Education (known as EuroPACE). The aim is to provide industrial Europe with up-to-date educational courses in six new-technology areas. But the project has also raised questions about the future linguistic balance of Europe.

EuroPACE, inspired by the US National Technological University which started broadcasting in 1983, has taken some two and a half years to develop and is one of the first major achievements of the European Economic Community's COMETT programme, designed to promote collaboration between industry and universities in new technology.

The initiative for the project was that of Hewlett Packard, the US instrument and computer manufacturer, which has experience of business television in the United States. (The company uses satellite broadcasts for training its own staff, and held its fourth annual teleconference on medical ultrasonics, spanning the United States and Europe, only last Saturday.)

Four other European sponsors (British Telecom, IBM Europe, Philips and Thomson) helped to provide start-up capital for the scheme, while COMETT enabled universities to participate in its design. In the past two years, other sponsors have joined from France, Denmark, Spain, Portugal, the Netherlands, Italy and Scandinavia but there is none, as yet, from West Germany.

Technically, the project relies on the European regional communications satellite called Eutelsat (ECS), launched in 1985 and owned by a consortium of telecommunications authorities. The management of EuroPACE broadcasting and networking rests with France Telecom International, which owns the French satellite Telecom 1 (also launched in 1985, but not used for EuroPACE).

The materials on offer are in the fields of artificial intelligence and expert systems, telecommunications, program engineering, microelectronics, technology management and advanced industrial engineering. The courses are prepared and presented by international panels from industry and higher education. They are available to companies paying a 'sponsorship fee' of 25,000 ECU (\$28,000), but at half that price to small businesses and universities. Users otherwise need only a parabolic receiving dish and an electronic mail account to use the facilities on offer, and to take advantage of the willingness of instructors to answer

questions.

Because the broadcasts are not encrypted, anybody with suitable equipment could receive the courses without having to pay a fee. But Hubert Curien, minister of research in the French government, says that for the sake of reaching as many people as possible for the smallest cost, the project will rely on the honesty of its audience for the time being.

Meanwhile, it seems to have been accepted that all broadcasts will be in English for the foreseeable future, which has implications for the face of technological Europe after the intended single common market comes into force in 1992. But whereas most science and technology students in Europe are used to reading English journals and textbooks, it is by no means certain that they will be equally skilled at following courses or formulating electronic mail questions in English.

The language problem is most acute in France, traditionally resentful of the increasing use of English neologisms as well as of 'linguistic imperialism' in science generally. Although EuroPACE may show that the position is softening, Curien says that it will be "absolutely essential" to encourage language learning. As multilingual broadcasting in Europe is still technically a long way off, EuroPACE may find that English courses are as germane to European competitiveness as its planned curriculum of high technology.

Peter Coles

Investigation into Tajikistan earthquake

London

PLANNERS responsible for the building of an irrigation canal near Shashora village in Tajikistan, the Tajik capital, could face criminal charges, a TASS report from Dushanbe suggests. The special commission now investigating the cause of last month's earthquake and landslide in the Gissar valley are investigating whether seepage from the canal led to a dangerous build-up of groundwaters, and are seeking to establish why the village, built in such a vulnerable area, was not provided with mud-traps. The commission, said TASS, "may view the construction of the canal as critical". Mud-flows are a major hazard in the mountainous regions of Soviet Central Asia and the Caucasus. In 1973, such a flow almost overwhelmed Alma-Ata, the capital of Kazakhstan, and in March 1987 Tajikistan suffered a major flood when a dam became clogged by mud and burst.

Vera Rich

Further fears of HIV in India

New Delhi

FEARS of human immunodeficiency virus (HIV) contamination in the Indian blood pool have been heightened with the chance discovery of HIV antibodies in a widely used product derived from donors' blood. The government last week ordered a Bombay-based company, manufacturer of a wide range of blood products, to stop production and marketing of its products and has launched an investigation.

The government was alerted after the detection of HIV antibodies in samples of anti-rhesus (Rh) immunoglobulin, a vaccine routinely given to Rh-negative mothers delivering Rh-positive babies. The contaminated batch of vaccines has been on the market since August 1988, but officials said they did not know how many ampoules had actually been sold. The entire stock of the vaccines has been withdrawn from the manufacturer's retail outlets throughout India.

The contamination would have gone unnoticed had it not been that one purchaser of the vaccine was a physician at the All-India Institute of Medical Sciences (AIIMS) who casually asked a colleague to test it before injecting into his wife. Examination of more samples showed they were all positive for HIV.

Coming so soon after the discovery of HIV infection in eight out of 94 haemophiliacs screened last month at the AIIMS, the incident has scared the medical community as well as the lay public. But the AIIMS authorities say that women who might have received the contaminated vaccine are unlikely to develop AIDS because the vaccine has only antibodies to HIV, not the antigen itself. The Indian Health Organisation, a voluntary group in Bombay, has challenged this view and has threatened to sue the Bombay manufacturer and the government for having exposed the public to the risk of AIDS.

Under a statutory requirement, both imported and indigenous blood products must be certified free from HIV. But the health ministry admits that it has so far applied this rule only to imported products because it lacks the machinery and the resources to inspect and certify local products. There is no law at present to compel the screening of blood donors.

In the aftermath of the latest incident, the health ministry has decided to step up its surveillance of blood products. It has also directed the Indian Council of Medical Research to investigate whether the procedures and technology used by the Indian manufacturers are adequate to ensure the safety of their products.

K.S.Jayaraman

Mutations and plagiarisms

SIR—In a paper¹ on the origin of mutants, my colleagues and I suggested, in the discussion, that reverse transcription of variant mRNA molecules is one way whereby cells might be able to generate novel, testable phenotypes before these become genotype. We felt that all our readers would know the source of that idea and so we gave no references.

The speculation was based, of course, on Temin's classical work with the retroviruses, which led him to propose the "provirus hypothesis"² and earned him a Nobel prize in 1975. He suggested that "in development of an organism, information transfer from DNA to RNA to DNA would allow variability"³. "There would be certain regions of the genome predetermined for this type of transfer"³ and proviruses "would be important in embryonic development, antibody formation, and memory"². "In extreme cases", he wrote, "one could imagine that a product of provirus evolution would infect the germ line, become integrated there, and thus also affect progeny organisms. Such a process could provide part of a mechanism for inheritance of some acquired characters"³.

Dr E. J. Steele of the University of Wollongong in Australia apparently believes that these ideas have somehow become his "intellectual property", and he has written to many people and com-

plained to the Australian press, saying that my colleagues and I have committed plagiarism because we did not refer to him⁴.

Steele wrote an interesting book in 1979 about the possibility of inheritance of acquired characteristics by way of retroviruses⁵, and shortly thereafter he published some experiments on the subject⁶. Later he wrote about the possibility of generating antibody variability by selective reverse transcription⁷.

As far as I can determine, Steele has never considered whether spontaneous mutations always arise at random, which was the topic of our paper; nor has he

referred to any of the classical papers on the subject. Certainly, nothing he has written had any influence on our work or on anything we put in our paper.

I am writing this letter to set the record straight. Let us hope that we can now all go back to work.

JOHN CAIRNS

Harvard School of Public Health,
Boston, Massachusetts 02115, USA

1. Cairns, J., Overbaugh, J. & Miller, S. *Nature* **335**, 142–145 (1988).
2. Temin, H. M. *Perspect. Biol. Med.* **14**, 11–26 (1970).
3. Temin, H. M. *J. natn. Cancer Inst.* **46**, ii–vii (1971).
4. *Nature* **337**, 101 (1988).
5. Steele, E. J. *Somatic Selection and Adaptive Evolution: On the Inheritance of Acquired Characters* (University of Chicago Press, 1979).
6. Gorczynski, R. M. & Steele, E. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2871–2875 (1980); *Nature* **289**, 678–681 (1981).
7. Steele, E. J. & Pollard, J. W. *Molec. Immunol.* **24**, 667–673 (1987).

Progress since 1900?

SIR—Your interest in educational matters reminds me that in 1900, when the average person had to learn many more skills than we need today, only 6 per cent of Americans graduated from high school — but the nation survived and even prospered.

Now, there are 60 million people associated with formal education in the United States, most of whom nicely illustrate John Holt's thesis in *How Children Fail* that "the major difference between the good student and the poor one is that the poor student forgets right away while the good one is careful to wait until after the examination".

Meanwhile, the United States has become the largest debtor nation in the history of the world, with numerous problems, oodles of expensive degrees, relatively little wisdom and millions of bored, anxious and resentful students. For the questioning and testing process almost guarantees that most feel quite inadequate long before their brains are fully developed and before their lives have hardly begun.

We could, of course, using inborn growth, creative, competence and curiosity drives, decide to help kids learn as opposed to 'Coop 'em, teach 'em, test 'em and flunk 'em'. But, with certain relaxed and productive exceptions, we usually seem to prefer a rigid, repressive and anxiety-promoting system where the 'failure' of the many gives the grade of 'A' to the few.

ROBERT E. KAY

PO Box C,
Paoli, Pennsylvania 19301, USA

Credit where it's due

SIR—A. G. E. Pearse (*Nature* **337**, 300; 1989) is absolutely correct to point out that the phrase *Omnis cellula e cellula* had been used before 1839. The lapse is even less forgivable as there is a discussion of the issue in the splendid biography of

Virchow published 30 years ago by my old teacher, the late Professor Erwin Ackerknecht. Whether Raspail's earlier use of the phrase qualifies him to be the "father of the cell theory" is another (and rather scholastic) matter. As J. V. Pickstone points out (*Dictionary of the History of Science*, article on 'Cell theory'), Schwann himself first used the phrase 'cell theory'. In our current state of knowledge, we should perhaps limit ourselves to the conclusion that Raspail is (probably) the father of the phrase *Omnis cellula e cellula*.

W. F. BYNUM

Wellcome Institute for the
History of Medicine,
183 Euston Road, London NW1 2BP, UK

Disclaimer

SIR—On behalf of the board of directors of Foundation 41, I wish to disassociate the foundation from the contents of a letter by Dr William McBride (*Nature* **336**, 614; 1988) and to say that it is unfortunate that the foundation's name and address were appended to that letter.

JOHN DARLING

Foundation 41,
365 Crown Street, Surry Hills,
NSW 2010, Australia

Error acknowledged

SIR—The writer of the article "Lamarck, Dr Steele and plagiarism" (*Nature* **337**, 101; 1989) seems to have committed what is known, I believe, in Great Britain as a school-boy howler. It would have been difficult for Copernicus to analyse Tycho Brahe's measurements as he died three years before Tycho was born. Obviously your writer confused Tycho with Johann Kepler.

FRANK W. DOBBS

College of Arts and Sciences,
Northeastern Illinois University,
Chicago, Illinois 60625-4699, USA

■ The presumption is sadly correct. —
Editor, *Nature*. ▶

Reactor lives

SIR—Our fast reactor is alive and well!

It is not correct to say, as Peter Coles did (*Nature* **337**, 197; 1989), that "Britain and West Germany have effectively abandoned their fast-breeder programmes".

In the United Kingdom we have assured funding for our fast reactor (PFR) until 1994 and for the fuel reprocessing plant on the same site until 1997. We also have a contribution of £10 million a year from the Department of Energy for our fast reactor research programme. This month we will sign up with the best Germans and the French, creating a European fast reactor programme in which we shall all collaborate. Fast reactors make much more efficient use of uranium than do thermal reactors and could supply the world's energy demand for a thousand years after we have burned up the coal, oil and gas. They do not contribute to acid rain or the greenhouse effect and are affordable. If successful, the European collaboration should lead to the construction of a series of fast reactor power stations.

J. H. GITTUS

UK Atomic Energy Authority,
11 Charles II Street,
London SW1Y 4QP, UK

Is there a God?

SIR—Bruce Denness (*Nature* 336, 614; 1988) suggests that God was smart enough to cover his tracks in his recent creation of the Universe by fabricating evidence of an ancient origin. That God should engage in such a fraudulent practice, which could be designed to mislead the most intelligent of his creatures, must undermine any trust we have in him. If God tells us lies, of what value is his Word?

If God is honest, the Universe is as old as the evidence suggests. If God doesn't exist at all, the Universe is as old as the evidence suggests.

PETER G. SWINDELLS

4 York Avenue,
Willenhall,
West Midlands, WV13 1JJ, UK

SIR—The argument put forward by Bruce Denness was proposed and developed at great length by the British naturalist Philip H. Gosse (1810–88) in his book *Omphalos* (c.1857).

Bertrand Russell in *Religion and Science* (London, 1958) made the following comments on this theory.

"There is no logical possibility of proving that this theory is untrue . . . but if once such possibilities are admitted, there is no reason to place the creation of the world at one point rather than another. We may have all come into existence five minutes ago, provided with ready-made memories, with holes in our socks and hair that needed cutting. But although this is a logical possibility, nobody can believe it; and Gosse found, to his bitter disappointment, that nobody could believe his logically admirable reconciliation of theology with the data of science".

JOHN BLEEKER

Ch. Des Etroubles 19,
1213 Onex, Switzerland

SIR—Bruce Denness has recently confessed that he "cannot find a way around" the argument that, "if God was smart enough to create The System, he was certainly smart enough to cover his tracks, that is he could have 'implanted' the geological [and] astronomical record so that what many of us now see as a scientifically pre-Creation history is merely a divine artefact." Let me reassure him, and any others similarly distressed, that they are neither the first, nor the most luminous, scientists to be flummoxed by this devious conjecture.

Such arguments are dismissed, not by supplying contradictory evidence, but by observing that there can be none. Such diabolical speculations simply define conflicting data out of existence. Science gives them short shrift, not because they are demonstrably wrong, but because they are scientifically vacuous: (1) they are unfalsifiable, because they fail to state how the

world we observe would be different if it were *not* a divine artefact; (2) they have no explanatory power, as they cower from the question of why *this* particular divine artefact, out of the infinite alternatives; (3) they are predictively useless, as the putative divinity of the Universe's origin does nothing to restrict its sphere of future possibilities; and (4) they are radically unparsimonious, because they impose an extraordinary burden of unverifiable assumption, while conferring no predictive utility in return. Science refuses to believe in miracles, because believing in miracles is scientifically useless, not because they can be scientifically refuted.

The fact that science has chosen falsifiability, parsimony and predictive and explanatory utility as its precepts, and therefore cannot offer proof or data for them, does not mean that, in Denness's words, "science would also appear to be a religion". Rather, it is precisely the adoption of these precepts (instead of others) that fundamentally distinguishes the scientific enterprise from the religious experience, and from most other aspects of daily life.

JAMES KIRCHNER

Energy and Resources Group,
University of California,
Berkeley, California 94720, USA

SIR—Bruce Denness's conclusion that God was smart enough both to create The System and to cover his tracks thereafter (*Nature* 336, 614; 1988) has scriptural backing. Isaiah 45.15 affirms: "Verily thou art a God that hidest thyself, O God of Israel, the Saviour." (King James version.)

In Job 38.4, God asks Job: "Where were you when I laid the foundations of the earth? Tell me if you know so much." (The Living Bible version.)

As for his second point that science would also appear to be a religion, at least one notable scientist in the person of Freeman Dyson supports him by arguing that science and religion are similar (D.G. King-Hele *Nature* 332, 748; 1988).

REGINALD T. CHELVAM

1740 West 27th Street, Suite 301,
Houston,
Texas 77008, USA

SIR—Bruce Denness asks for a way around the argument about a creator-god who was "smart enough to cover his tracks" by implanting geological and astronomical evidence for pre-creation history.

Besides the question of what might be the motivation of this puckish god, the trouble with Denness's idea is the same as that with the slightly more extreme idea that the world began literally one minute ago — including, of course, all the books in the libraries and all the memories in our minds. The trouble is twofold: (1) By the very construction of such hypotheses, it is

impossible either to refute or to confirm them, and (2) even worse is the fact that they are sterile: absolutely nothing logical or practical follows from them. So although such ideas make some people emotionally comfortable, they are simple dead ends, and not any kind of science.

To be sure, various religions add on to the supernatural-creation hypothesis several more hypotheses, such as the Garden of Eden, Noachian flood, Ten Commandments, Virgin Birth and Resurrection. But these are just additional hypotheses, not logically connected with one another, and the truth or falsity of one of them says nothing about the truth or falsity of any other. For example, if someone should ever find a piece of very old wood on Mount Ararat, this could be construed as evidence for a great flood, but it would mean nothing at all to the hypothesis of supernatural creation.

It needs to be pointed out over and over again that supernatural creation is not a theory, but just a hypothesis (and a sterile one at that), while biological, geological and astronomical evolution are full-blown scientific theories, consistent with each other and with a lot of data.

EDGAR PEARLSTEIN

Behlen Laboratory of Physics,
University of Nebraska,
Lincoln,
Nebraska, 68588-0111, USA

SIR—The answer to Bruce Denness was surely given many years ago by Albert Einstein: "*Raffiniert ist der Herr Gott, aber boshaft ist Er nicht.*" ("God is cunning but he's not malevolent.")

NICHOLAS RYAN

Nicholas Ryan Consultancy Ltd,
14 Charlotte Mews,
London W1P 1LN,
UK

Capital idea

SIR—In your leading article "Student debtors rally" (*Nature* 336, 411; 1988), you say "Capital cities are incomplete without occasional conflicts between students and the authorities . . .".

This is the very reason why the wise authorities of the past would never have thought of having a university in their capital city. Thus Padua not Venice, Pavia not Milan, Uppsala not Stockholm and, of course, Oxford and Cambridge not London.

The fact that Venice, Milan, Stockholm and London now all have universities of a sort does not mean that some authority might not seriously consider reverting to the old wise practice.

MARCO FRACCARO

Collegio Cairoli,
Piazza Collegio Cairoli 1,
27100 Pavia,
Italy

Fateful discovery almost forgotten

The discovery of the fission of uranium exactly half a century ago is at risk of passing unremarked because of the general ambivalence towards the consequences of this development. Can that be wise?

JUST over 13 years from now, there will be great celebrations on the fiftieth anniversary of what will then be recognized — as it is, of course, already — to have been one of the great discoveries in the history of science, that of a plausible structure for the nucleic acid polymer called DNA from which has flowed almost the whole of what is now called biology.

Those of an anxious or paranoid disposition should be prepared soon to reserve places at the celebratory symposia that will be held around Easter 2002. They and the rest of us should meanwhile reflect that this year, 1989, is also the fiftieth anniversary of a discovery which has done as much to change the world we live in, and as quickly, as even the discovery of the role of DNA in genetics. Yet the crowds are not fighting for places at the celebratory symposia — indeed, there are hardly any symposia to attend.

Yet nuclear fission is just half a century old. For more than half a decade, Fermi (at Rome) had been gathering data on the consequences of bombarding uranium nuclei (atomic number 92) with neutrons. Artificially radioactive materials abounded in the products. It was natural that Fermi should suppose that neutrons would first be absorbed by the heavy nuclei and should then decay (losing an electron), leaving the plausibly radioactive nucleus of some literally artificial element, with atomic number greater than 92. But there were implausibly many radioactive species to account for.

The key to this puzzle came only at the end of November 1938, when Otto Hahn and F. Strassmann, from the Kaiser Wilhelm Institute in Berlin, announced that Fermi's disconcerting artificial radioactivity included at least one isotope of barium, with atomic number 56 (*Naturwissenschaften* 26, 756; 1938). The general expectation that nuclear reactions of this kind would effect only small changes in the atomic number, or the unit electric charge, of a nucleus was overturned. Physics was never again to be the same.

The explanation came just a few weeks later, on 11 February 1939, in a letter to the editor of *Nature* from Lise Meitner and her nephew, Otto R. Frisch (*Nature* 143, 239; 1939). In a simple yet sophisticated argument, they showed that neutron bombardment of uranium nuclei must induce their fission into smaller fragments. What followed — the Manhattan Project, Hiroshima and Nagasaki, Shippingport

NATURE No. 3615, FEB. 11, 1939

News and Views

A New Type of Nuclear Reaction

ON p. 251 of this issue of *NATURE*, an account is given of recent investigations by Prof. O. Hahn, Prof. L. Meitner and F. Strassmann on the bombardment of uranium by neutrons. Prof. Meitner and Dr. O. R. Frisch have discussed the development and implications of these results in a letter which also appears in this issue (p. 239). Experimental confirmation of these conclusions, it is claimed, is announced in the following cable, dated February 3, received from R. D. Fowler and R. W. Dodson, of the Chemical Laboratory, Johns Hopkins University. "We have bombarded uranium nitrate in a three millimetre ionization chamber with deuterium neutrons and found that particles causing a very intense ionization, at least five times that from natural uranium alpha particles, are produced. Fast neutrons from one milliamperes two hundred and fifty kilovolt deuterons produced thirty-five particles per minute. Placing paraffin around ionization chamber increases this to seventy counts per minute. We believe this to be confirmation of work of Hahn and Strassmann (*Naturwissenschaften*, Jan. 6; Frisch and Meitner *NATURE* [Feb. 11, p. 239—Editor]) in which activity ascribed to barium was found after neutron bombardment, and that these particles are barium ions of about one hundred million volts energy."

"We have also bombarded thorium oxide in similar ionization chamber, with deuterium neutrons. Fast neutrons in same intensity as above produce thirty intensely ionizing particles per minute. Paraffin does not increase the effect. We believe thorium is also disintegrated by fast neutrons into fragments of about one half thorium mass, having energies of about one hundred million volts."

(if anybody remembers that) and even Chernobyl — is inextricable from our contemporary history. It is also relevant that Meitner, previously at the Kaiser Wilhelm Institute but then evading the Nazis, gave her address as the Academy of Sciences, Stockholm. Her nephew, also Jewish, had made an earlier escape and had spent a year in London before settling with Bohr at Copenhagen.

Both the discovery of fission and its exploitation are, of course, heroic tales. At various times in the past half-century, they have been celebrated as such. But with the passage of the years, and for reasons which are thoroughly familiar, fission has become a dirty word. There are many people, even in research, who would prefer that history had taken a different course. For some, the ideal might have been that the Universe had evolved so that some fundamental constant, Planck's constant for example, had been sufficiently different from what it is as to have made the fission of uranium irrelevant. But that is not what the world is like, so people keep silent instead, hoping that 1989 will pass off quietly.

That view scorns the truth that the dis-

covery and development of fission remains one of the most enthralling demonstrations that ingenuity can be a match for nature, and is thus a disservice to the reputation of science. The discovery of fission was also literally an heroic endeavour whose importance can with hindsight be diminished only by disservice to the reputations of a band of mostly admirable people, many of whom are still alive. To fail to acknowledge that is also to fail to acknowledge that fission quickly became the chief agenda item in the relationship between science and government from which, on balance, both parties have benefited. But were there not Hiroshima and Nagasaki, four decades of fission-provoked cold war and then Chernobyl? May not seemly neglect be the best mark of this uncomfortable anniversary? Only if it is supposed that uncomfortable truths are best forgotten. That supposes that some history is so painful that nothing can be learned from it, which is a disservice to the intellect.

The ingenuity that fission has always commanded is already plain in the article from Meitner and Frisch, which is far more subtle than a mere jumping to the correct conclusion. Their objective is to show that uranium fission is the preferred interpretation of Fermi's data.

First, they show fission to be plausible, using Bohr's theory of the nucleus as a liquid drop held together by its surface tension, which nevertheless diminishes steadily as the charge (or the atomic number) increases. They discard the notion that smaller fragments might be formed from uranium nuclei by the quantum tunnelling of particles much larger than the helium nuclei that make α -particles, concluding that the "whole 'fission' process can thus be described in an essentially classical way". Even so, it is still astonishing that their estimate of the energy released in the fission of a uranium nucleus should have been 200 MeV, within a few per cent of the measured value.

Meitner and Frisch then show that fission makes better sense of the data accumulated since Fermi's earliest experiments than any other explanation. The paper is crammed with intuitive suggestions of radioactive β -decay chains for which to look among the light elements — for example, strontium-yttrium-zirconium. But they allow Fermi one transuranic element, supposed to be uranium-

239, but in reality plutonium-239 (of which much more was to be learned).

Even by present standards, the excitement generated by the developments is remarkable. Within a week, Frisch was reporting (*Nature* **143**, 276; 1939) the direct measurement of ionization caused by fission fragments and an estimate of the energy of the fragments in excess of 50 MeV. The two papers carry the same date, but one was delayed a week. Did *Nature* take the view that its readers could stomach only so much surprise? The following week (*Nature* **143**, 330; 1939), Bohr himself blessed Frisch's argument that fission could be accurately described by the fission of an oscillating liquid drop, dealing in passing with the question why such unstable nuclei did not disintegrate spontaneously (spontaneous fission had

towards ... exo-energetic transmutation chains is evident. However, in order to establish such a chain, more than one neutron must be produced for each neutron absorbed. This seems to be the case...." By April (*Nature* **143**, 680; 1939) the same group had estimated that the neutron-catalysed fission of a single uranium nucleus would yield an average of 3.5 neutrons (one to replace the catalyst and 2.5 for other purposes) and, by 3 June (*Nature* **143**, 939; 1939), that some of the spare neutrons liberated in fission carry energy in excess of 2 MeV. The stage was set for putting chain reactions to work.

That so much was accomplished so quickly is both a tribute to the ingenuity of those concerned and a sign that they understood the importance of the problem. Long before the summer of 1939,

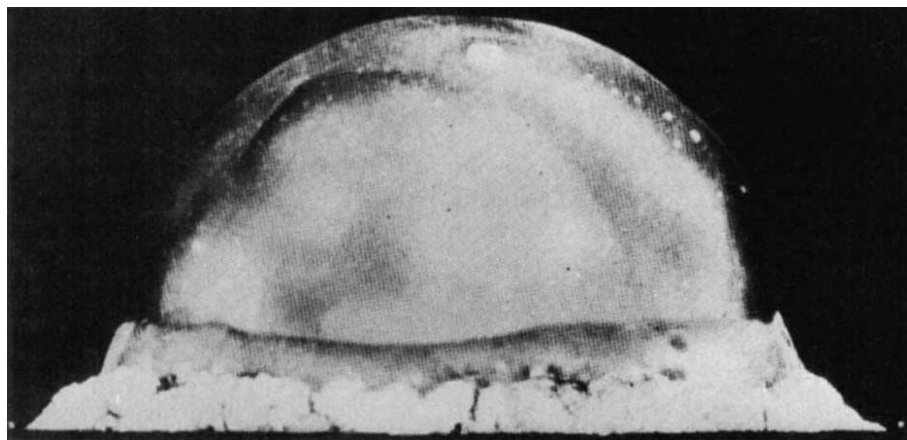
Frisch was about to put his manuscripts in the mail to *Nature*.

By the spring, the character of the phenomenon of fission had been independently confirmed. Bohr seems to have been the first to raise the crucial question of which of the two predominant isotopes of uranium (235 and 238) is chiefly responsible for fission by slow neutrons. He told the spring meeting of the American Physical Society at the end of April that uranium-235, if separated from the larger bulk of uranium-238, would sustain a chain reaction, thereby starting the argument about the feasibility of separating these isotopes which lasted more or less until the separation plant at Oak Ridge had been commissioned.

Frisch, by then at Birmingham with Rudolf Peierls and the Australian Mark Oliphant, seems to have been the first to recognize that, while the fission of uranium-235 by slow neutrons might allow the construction of a nuclear reactor, only the fission of the same isotope by fast neutrons would make a bomb. He and Peierls persuaded Oliphant of that, and were eventually allowed to persuade F. Simon to divert his attention to the separation of the isotopes of uranium by diffusion. By the summer of 1940, the British government had been convinced that what was called a "super-bomb" would be feasible and, with other things on its mind, took to persuading the United States that such an effort would be in the mutual interest.

By the summer of 1940, the other route to nuclear weapons had been signposted in the United States by the discovery at Berkeley of the element with atomic number 93 (now called neptunium) by E. McMillan and Philip Abelson (until recently the editor of the journal *Science*). This proof that trans-uranic elements did indeed exist seems quickly to have put into people's minds the certainty that some of them might be fissile and might therefore provide an easier means of producing materials for bombs and reactors. It was a matter of weeks before plutonium made its appearance.

Hindsight brings many benefits, but also the disadvantage of making the process of discovery seem a commonplace, the stuff with which the introductory pages of text-books are filled. It is not easy now accurately to reconstruct the state of innocence of those who so quickly comprehended the ramifications of the discovery of nuclear fission. The fact of the neutron was a few years old (and a neutron source was a speck of radium mixed with beryllium). That there are forces between nucleons (neutrons and protons) was phenomenologically established, but Yukawa's guess that they are mediated by mesons would not be established until after the Second World War. But hopes that the motion of nucleons within nuclei would be analogous with those of elec-



The first atomic explosion, the Trinity Test, at Alamogordo in the New Mexico desert on 16 July 1945.

not then been measured) by noting that the latent heat of condensation of a neutron onto a uranium nuclear-drop was necessary to set the nucleus in oscillation.

Fission became more serious on 18 March (*Nature* **143**, 470; 1939) when H. von Halban, F. Joliot and Leo Kowarski at the Collège de France described the experiments by which they had shown that spare neutrons are released in the fission process. The starting point was the observation that uranium nuclei contain as many as a dozen neutrons above the requirements of stable nuclei of half the mass, that some of the extra neutrons are incorporated in the fission fragments (which is why their radioactivity is predominantly β -activity) but that some may be liberated in the fission process.

The experiment consisted simply of the measurement of the diffusion of slow neutrons through aqueous solutions of uranyl nitrate and ammonium nitrate (used as an inert control). The observation was that neutrons diffuse further into uranyl nitrate, for which the only explanation is that those absorbed as triggers of individual fission events are more than replaced by neutrons released by the fission events.

The authors concluded that the "interest of the phenomenon ... as a step

people in the know in the United States and Europe were openly speculating about bombs, nuclear reactors and perhaps both. It is remarkable that, at this point, with the outbreak of the Second World War in Europe just a few weeks ahead, the Royal Society should have asked Otto Hahn to London to describe his work in public on 23 June, and that Hahn should have accepted. (*Nature* paid Dr Norman Feather as he then was 10 shillings for his report of the occasion at **144**, 46; 1939.) Within a year of the appearance of the paper by Hahn and Strassmann, more than 100 others had appeared.

The pace of these developments was too great for the transatlantic communications of the times — the telegraph cable and the ocean-going liner. During the first half of 1939, the European work on fission appeared almost exclusively in *Nature*, that in the United States in *Physical Review*, but inter-communication depended largely on chance.

Thus the content of Frisch's paper with Meitner, and of his demonstration of energetic nuclear fragments from the fission of uranium, reached the United States by means of Bohr, who happened to be leaving Denmark for the United States at the point in mid-January at which



Historic sight: Robert Oppenheimer and General Leslie Groves after the Alamogordo test.

trons in atoms, and thus described by simple quantum numbers, had already been disappointed. And while the experimentalists had accumulated several years of experience in the study of nuclear reactions of a familiar character, fission required endless ingenious improvisation. To have planned to recognize radioactive species by the characteristic energy of their γ -rays would have been to anticipate the future.

It is remarkable, and a comment on the way in which the research profession's judgement of its own recent past is compromised by ambivalence towards the consequences, that this achievement is now more often described as "fateful" than as "great". But can it in the long run be wise that such a tale of cleverness in the face of a natural conundrum should be neglected because some of the consequences are uncomfortable? It is as if an applicant for a job were to suppress from his *curriculum vitae* a spell of employment as an income-tax collector, whose practitioners are not always popular.

This neglect is also a disservice to those involved, whose influence then and since has done much to shape the framework within which the profession of science must now work, not only in the West but also in the Soviet Union. Much has rightly been made of the circumstance that refugees from Germany and its occupied territory were conspicuous in the understanding of fission and what followed. Can there ever have been another intellectual migration on such a scale? Meitner, an Austrian Jewess, left Berlin too late for

comfort. Britain took (and kept), for example, Simon, Peierls and Mendelssohn, and the United States, the galaxy of talent of Einstein, J. von Neumann, E. P. Wigner, E. Fermi, E. Teller (who had stopped off in London on the way), Weisskopf and a host of others.

The scale of the migration is naturally impressive, but the ease with which it was assimilated is even more so. People newly arrived from hostile states might have expected to spend time in camps for aliens (which is why the Fermi family buried part of Enrico Fermi's Nobel prize money in cash in the garden of their New York house), but in the event were welcomed as urgently needed colleagues. It might have been different if the study of nuclear reactions had been destined to remain an academic subject. But in the event, the migrants played a role going far beyond their narrow expertise.

The immigrants from Europe were the first to make and then to sustain in the face of bureaucratic indifference the argument that the search for a practicable nuclear explosive was a necessary precaution against the possibility that Germany might do the same successfully. Hungarian Leo Szilard, an enthusiastic fixer of a physicist (with a patent of dubious value on energy-generation by nuclear transmutation up his sleeve) was among the first in the United States to urge voluntary restraint on the publication of research on fission. In the end, it fell to immigrant Einstein to write the two letters that persuaded President Franklin D. Roosevelt to launch the Manhattan Project. Even with hindsight, nobody can seriously pretend that the argument was incorrect.

The influence of the hectic months following the discovery of fission on the sociology of the profession of science has been profound. Practical needs led quickly to the practice of what had previously been a precept only — that a person's value as a scientist rests not on age, or background, or even paper qualifications, but on the understanding in his head which he can communicate or the skill in his fingers which he can put to use.

Upstart students who now take as a given their right to function within their sphere of competence as independent scientists must thank for their privileges the utilitarian tradition springing from the hunt to put fission to military use that the value of an idea or a technique can be judged independently of its provenance.

Friendliness is another legacy of those times. Those still alive who were at Los Alamos before 1945 still recall their time there with affection. The proof that it is possible to practise physics in shorts and shirtsleeves was plainly influential. So, too, was the practice of eating lunch from brown paper sacks. High-energy physics laboratories the world over seem to strive to keep these outward traditions alive.

But the more durable cement of these relationships is the sense that people appear still to enjoy that they had been given a licence by a government frequently unconvinced to turn a wild idea into a reality and in the process to demonstrate that even abstract science can be useful.

Los Alamos was not of course the only important laboratory of the Manhattan Project (there were also Berkeley, Chicago and the production plants, as well as the British-Canadian enterprise at Chalk River) and nuclear energy was not the only important vehicle by which science made a wartime contribution to military affairs. (Radar and cryptography seem also to have been powerful wellsprings of comradeship.) But it is remarkable that people's affections for Los Alamos appear to have survived even the unfriendly traumas of the post-war period. The unpleasant procedure by which Oppenheimer's security clearance was withdrawn by the Atomic Energy Commission in 1956 has scarred all the participants and many others. By contrast, it is odd that the revelation, at about the same time, that Klaus Fuchs had been disloyal to the Los Alamos cause seems not to have filled those who worked there with the chagrin that might be expected of people betrayed.

But the Manhattan Project was different partly for its scale and, eventually, for its secrecy, and partly because of its daring — the risk of failure must always have seemed uncomfortably great, too many of the calculations might have been made on the backs of envelopes and too many of the careful calculations used numbers that were measured only crudely. Those who enjoy the game of rewriting history should perhaps ask what would have happened if the Manhattan Project had failed. How long the Pacific War would have dragged on after 1945 is the obvious but unanswerable question. That of what the reputation of science would have been is not nearly as difficult to answer. At one stage in the argument between the people at Los Alamos advocating that the two manufactured bombs should not be used in anger but merely demonstrated, the bureaucracy seriously countered that the US Congress would "want to know what happened to all that money".

In the event, science emerged from this adventure with its reputation enormously enhanced. That is not surprising — and, some will say, not particularly creditable either. But the case is not as simple as it may seem after nearly half a century.

Despite the conflicts of the intervening decades, it is difficult now to reconstruct the urgency of the general conviction at the discovery of fission that the world was heading for a war in which the stake was no less than the preservation of the fabric of civilized society.

The energy and flair with which the Third Reich had rebuilt the German Navy



Lise Meitner, who, with Otto Frisch, first explained the nuclear fission process in 1939.

and created from scratch an aircraft industry lent weight to the fear that it might also be able to make nuclear explosives more quickly than its adversaries. In present circumstances, it is possible to contemplate and even to advocate agreements between potential adversaries that some kinds of weapons systems should be deliberately left on the shelf, but the world knows the difficulties of those projects. The case for building a bomb was irresistible once the feasibility of the project had been demonstrated. The wisdom of the first use, against Hiroshima and Nagasaki, is another matter.

But should not the scientific community have been more alert to the consequences of its adventure? There are two apparently contradictory responses. First, although the explosive power of the first bombs and the effects of radiation and radioactivity on civilian populations appear to have been estimated in advance with reasonable accuracy, and enthusiasm had developed for the promise of endless and cheap electricity from nuclear reactors, nobody had been able to calculate what the political consequences would be. (But P.M.S. Blackett, in Britain, accurately predicted nuclear stalemate between the United States and the Soviet Union and, after a distinguished wartime career in operational research, was for some years afterwards ostracized by the British government for the book in which he said as much.)

Second, however, Los Alamos and Chicago between them produced not only the first nuclear weapons but the tradition that people in research have an obligation towards the use made of their creations. The second thoughts at two laboratories in the early months of 1945 about the proper use of their nuclear explosives were ineffectual, for there was no common view of what might be done with them except

what the military were planning, and ineffective, for the argument with the military was never joined. Similarly, the hope nurtured by many that nuclear weapons were qualitatively novel and thus must serve as catalysts of novel kinds of international understandings had been discovered to be wishful thinking by 1947, when the Baruch plan to internationalize nuclear energy inevitably collapsed, partly because of its internal contradictions. Yet the tradition that science need not be neutral on the uses people wish to make of it has flourished fitfully in the intervening decades. If it fades, or is entirely overtaken by the notion that what makes money must be worthwhile, the blame will attach not to the Manhattan Project people but to their successors.

The reputation of science and its relationship with government is another matter. In the United States, the bombs helped to win a place in government for science and scientists. Until the Nixon administration, there was a President's Science Advisory Committee, one of the means by which the federal government was persuaded to shoulder responsibility for the support of research in general, not simply in public health and agriculture. To suggest that the present output and quality of science in the United States is attributable to the Manhattan Project would be a travesty of the truth, but success helped. While the comparison is remote, much the same appears to have happened in the Soviet Union. The relative independence of the Soviet Academy of Science between 1941 and 1986 owes much to the power of the Kapitza, Kurchatovs and Artsimovitchs who carried through the Soviet nuclear weapons project.

The other side of this coin is naturally less endearing. The intimacy of science and government in the Manhattan Project has left government with the impression that science is properly an instrument of national policy. The rights and wrongs of that position are rarely defined with the clarity they rightfully demand. It is, of course, proper that a government seeking to carry through some worthwhile technical project should expect that its nationals who happen also to be scientists will sign up for the enterprise, but there is a worrying temptation to suppose that those who might help have no particular right to hold opinions about the wisdom of the project. People's devotion to the Manhattan Project has led governments to believe that it should be possible to whistle up a crowd of willing workers even for projects whose rationale is much less compelling than, in the early 1940s, was the project to make nuclear weapons. The precedent of the Manhattan Project is awkward in relation to the Strategic Defense Initiative, for example.

It must also be confessed that the success of the Manhattan Project, or the

demonstration it provided that the most difficult tasks can be accomplished by sufficient zeal and daring, went to people's heads. Sadly, the consequences are mostly the well-known flaws in the what seemed at the beginning to be the benign face of fission — the use of self-sustaining chain reactions for generating electricity.

In 1955, for example, the British government announced its plan for a dozen nuclear reactors whose construction would begin within a decade and which would span the whole gamut of conceivable designs, from thermal gas-cooled reactors to liquid-sodium-cooled fast reactors. That same year, the first UN Atoms for Peace conference at Geneva was told of how the running cost of generating electricity would be so small that consumers might not have to pay for the number of kilowatt-hours used. The Manhattan Project was the model on which these dreams evolved, but could not but be misleading: making a few tens of kilograms of fissile material, difficult though it had been, was a simpler task than creating a whole new industry. Los Alamos had necessarily little to say about the demands of the long haul — durability and reliability in particular.

But these are not usually the grounds on which the nuclear industry is now commonly regarded as a second malign face of fission. Instead there is the fear of nuclear accidents (of which Chernobyl is the worst so far), of the routine release of radioactive materials from reactors and reprocessing plants and the complaint that there is no sure method for the long-term disposal of nuclear waste. Yet, ironically, these problems have been anticipated more accurately than those of building and operating nuclear reactors economically. It is natural that nuclear engineers should nurse a sense of grievance at the unfairness of these charges, even though their own over-confidence has provoked the distrust they accurately sense.

The truth is that the nuclear industry has everywhere — Chernobyl notwithstanding, even in the Soviet Union — been more diligent about the protection of people from radiation and its consequences than in regard to the engineering and economy of its plants. The view now gaining ground that the risk of further nuclear accidents cannot be contained acceptably is tantamount to the assertion that nuclear technology is unmanageable at the standards it is reasonable to expect of it. If that is how the argument about the future of nuclear power is resolved, the result will be that a potentially valuable technology must be left on the shelf even at a time when the threat of climatic change makes nuclear power particularly valuable. That would be an ironical way of celebrating the half-century of the discovery of fission — and a unique confession of technical incompetence. □

Research, misconduct and Congress

A private conference ten days ago seems chiefly to have demonstrated that there is some way to go before researchers learn to live with — or placate — their lawmakers.

Washington

A CONFERENCE ten days ago at the Banbury Center at the Cold Spring Harbor Laboratory, with the curious title "The Ethos of Scientific Research", should probably be considered a valuable milestone showing how far the present debate about issues of scientific misconduct has come. Scientists, members of congressional staffs, a few journal editors and others who have contributed ideas for coping with the misconduct problem spent a day and a half in a meeting that seems to have been informative, but punctuated with episodic incredulity as scientists and legislative staffs alike learned how divergent their views can be. That is why cynics may choose to observe that this milestone also shows how far the debate still has to go.

The very mention of the word misconduct in scientific or congressional circles these days generates a hysterical craving for confidentiality by insiders to the process. But keeping a lid on such controversial issues is virtually impossible, and although the Banbury meeting was closed to the press, participants were willing, and in some cases anxious, to discuss their impressions afterwards.

The scientists in attendance were a stellar lot: James Watson, director of the Cold Spring Harbor Laboratory, Harvard's Walter Gilbert and Mark Ptashne, Rockefeller's Norton Zinder and Eric Kandel and Richard Axel from Columbia, to name a few. The names from Congress were less familiar, as only staff and no congressmen themselves attended, prompting one scientist to grumble that "we send them our first team, and they don't send theirs".

If there was ever hope that a consensus would emerge on handling scientific misconduct, that hope was not realized. What is agreed is simply that there is a problem whose scale or even seriousness is unknown. Although most research organizations and federal agencies would maintain that the number of cases of misconduct is small, there is little empirical evidence to back that up. But the reality now is that the size of the problem is no longer a crucial question; people in Congress see a problem, so there is one. Even if the scientific community could respond with carefully planned studies proving that misconduct is a negligible problem, until the political winds change it would not make much difference.

The easy approach, and one which would in general be pleasing to researchers, goes something like this. Congress has rightly identified a problem in the practice of science, the scientific community is appropriately grateful and will now take care of it if it is left alone.

But Congress has, with good reason, become wary of petitioners making "trust me" arguments. Washington is full of special pleaders on behalf of industries, agriculture and the armed services, to list but a few. And dealing with misconduct is an unpleasant task for all concerned, as the recent investigation of the *Cell* paper (see page 490) proves. No institution has shown much aptitude for conducting these investigations, and the National Institutes of Health (NIH) are no exception. The issues to be decided are not in question: a mechanism must be developed to determine when charges deserve investigation, to protect the rights of both accuser and accused, to carry out the investigation in a thorough and speedy way and to determine what punishment should be dealt to those found guilty.

The American Association of Universities has proposed a framework for such investigations that universities would in the first instance undertake (see *Nature* 337, 196; 19 January 1989), but there is a legitimate fear that universities would lose some of their collegiate character if they were assigned the role of policeman for their own research communities.

According to the Health Research Extension Act of 1985, those applying for federal research funds must guarantee that their institutions have mechanisms to deal with accusation of misconduct, but a report prepared by the Office of the Inspector General of the Department of Health and Human Services found that only 22 per cent of institutions receiving NIH grants have such procedures in place.

The Public Health Service has also served notice that it is considering new regulations that would spell out the responsibilities of institutions receiving federal research money. In an advance notice of proposed rulemaking in the 19 September issue of the *Federal Register*, the service raised the possibility of an office of scientific integrity to review all charges of misconduct and to determine what action would be most appropriate. A similar idea was incorporated into proposed legislation that circulated at the end of the last congressional session.

Although the legislation was never formally proposed last year, it is an active subject for discussion once again this year.

The need for the scientific community to work with Congress is at present complicated by poor relationships between scientists and congressional staffs, especially those staff members working for investigation and oversight subcommittees. The role of these committees is not to draft legislation or to appropriate money, but to ensure that laws are being implemented as Congress intended, and that money is being spent in a reasonable and appropriate fashion. But the Banbury conference — with the exception of Watson and Gilbert who communicated genuine concern for the issue — seems to have been marked either by arrogance or hostility or both on the part of some researchers towards these staffers, which does the former little credit.

Walter Stewart, who has temporarily left his NIH laboratory to work for the House of Representatives Energy and Commerce investigation subcommittee, came in for a particularly hard time from the researchers present. Stewart has incurred researchers' wrath for his investigations of alleged scientific fraud, investigations that have been marked at times by an almost religious fervour. Indeed, at the Banbury meeting, Stewart astounded participants by equating the moral taint of scientific fraud with that of the Holocaust. Although his point was that responsibility for identifying and tackling problems falls on everyone's shoulders, the idea that an incorrect scientific paper, even one written with knowing deception, can be in any way compared with the slaughter of 6 million people suggests that his enthusiasm for his work has exceeded reasonable bounds; he may no longer be a credible force in these investigations.

Congress has already served notice that it will continue to hold oversight investigations on this issue. And the scope of its interest appears to be broadening. In addition to the legislation that may emerge on misconduct as such, there may also, for example, be new laws explicitly forbidding researchers from accepting federal money to work on projects that may benefit companies in which they have a financial interest. The principle seems to be that, as spending on science grows, Congress will take an increasingly close look at how that money is spent — and by whom.

Joseph Palca

G-protein subunits

Who carries what message?

Henry R. Bourne

A SMOULDERING 2-year-old controversy over how G proteins work has finally begun to generate more light than heat. The initial question referred to a specific signalling pathway in plasma membranes of heart cells: which subunit of G_i , a member of the ubiquitous family of heterotrimeric G proteins, couples muscarinic acetylcholine receptors to the opening of potassium channels^{1,2}? Is it the GTP-binding α -subunit or the membrane-attached $\beta\gamma$ -complex? Not surprisingly, one answer turns out to be right, the other wrong. In retrospect, however, the wrong answer provided a clue to the potential existence of a novel signalling pathway mediated by G proteins. Recent studies³⁻⁶, including two reported on pages 557 and 555 of this issue^{7,8}, reveal the broad outlines of such a pathway in mammals, sea slugs and baker's yeast. The outcome is more interesting than originally imagined by those on either side of the controversy.

The initial dispute grew out of experiments in which purified α - or $\beta\gamma$ -subunits of G proteins were added to the cytoplasmic side of patches of plasma membrane plucked from heart cells. Birnbaumer and Brown's group¹ reported that the active form of the α -chain of G_i (α_i bound to the GTP analogue GTP- γ -S) opens K^+ channels in the patch. Clapham and Neer's group² reported the opposite result: $\beta\gamma$ -but not α -subunits open K^+ channels. Which group is right? Before revealing the answer, let us consider the question in a broader context.

The orthodox view of G-protein signalling^{9,10} points to α_i as the mediator, in accordance with precedent: the α -chains of G_i and transducin mediate, respectively, stimulation of adenylyl cyclase by hormone and stimulation of retinal cyclic GMP phosphodiesterase by light^{9,10}. Although other systems are not understood in the same degree of biochemical detail, the numerous additional effector targets of G proteins (phospholipase C, phospholipase A₂, K^+ channels and Ca²⁺ channels, for example) seem to require corresponding structural diversity within the class of G-protein components that regulate effectors. Such diversity could be supplied more plausibly by α -chains, of which nine distinct primary structures have been identified, than by β (two) or γ (two or three).

Although conventional wisdom has favoured α -chains, the apparently heterodox¹¹ notion that $\beta\gamma$ can act independently of α is nonetheless perfectly consistent with our understanding^{9,10,12} of

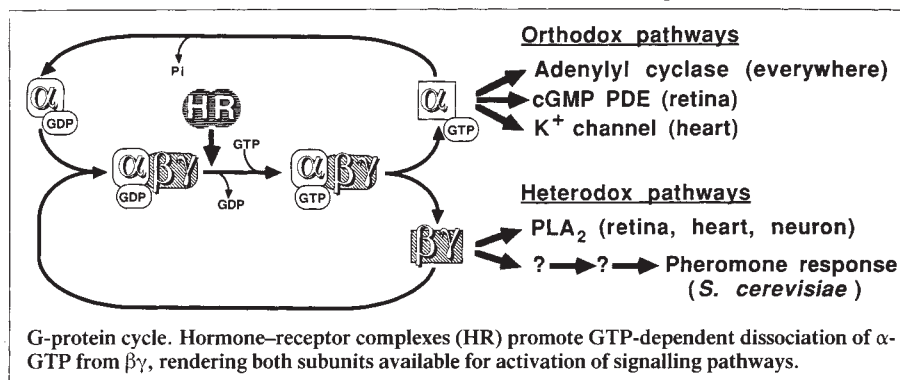
the biochemical cycle involved in receptor-effector coupling by G proteins (see figure). Hormone-receptor complexes activate G proteins by catalysing replacement of bound GDP by GTP. Binding of GTP to the G-protein α -subunit causes it to dissociate from $\beta\gamma$, thereby generating two potentially active subunits, α -GTP and $\beta\gamma$. Eventually, hydrolysis of the bound GTP returns α to its inactive GDP-bound form, which binds $\beta\gamma$ with high affinity; until then, at least in principle, each subunit remains available for interaction with an effector enzyme or ion channel.

This more ecumenical view is supported by experiment. For example, the $\beta\gamma$

mammalian cells^{3,4,6}.

Most surprisingly, at least to an α -chain chauvinist, genetic deletion of the α -chain gene does not prevent mating, but instead produces the opposite phenotype — a cell in which the pheromone-response programme is always turned on^{3,4}. In contrast, as described in a report⁶ in the current issue of *Cell*, deletion of either the β - or the γ -gene abolishes a cell's ability to respond to pheromone; such deletions also block the constitutive response pathway caused by genetic deletion of α . Taken together, these results indicate that β and γ , but not α , carry the message from receptors to effector elements of the pheromone response. This genetic evidence⁶ also suggests, as predicted earlier³, that the GDP-bound form of α suppresses activity of β and γ , and that the pheromone-receptor complex activates the response by promoting release of $\beta\gamma$ from α -GTP (see figure).

What about regulation of cardiac K^+



(but not the α) subunit of transducin stimulates phospholipase A₂ (PLA₂) activity in retinal rod outer segments¹³. In other tissues, the chief role of PLA₂ is to catalyse cleavage of membrane phospholipids to produce arachidonic acid, the precursor of prostaglandins, leukotrienes and other lipid messengers. The physiological relevance of lipid metabolites released by the action of $\beta\gamma$ on PLA₂ is unknown.

But α -chain chauvinism has received a more damaging blow from an unexpected quarter — convincing genetic evidence^{3,4,6} that the β - and γ -chains of a yeast G protein, but not the α -chain, are required for response to mating pheromones. In *Saccharomyces cerevisiae*, specific peptide pheromones produced by cells of one mating type initiate in cells of the opposite mating type a programmed response that prepares them for conjugation. The pheromones associate with receptors that are structurally similar to the receptors that talk to G proteins in mammals¹⁴; in turn, the pheromone receptors act through a G protein whose subunits have been identified genetically, rather than biochemically; primary structures of its α , β and γ subunits closely resemble those of the cognate G-protein subunits in

channels? After the first pair of contradictory reports^{1,2}, the Clapham/Neer group found that preparations of both $\beta\gamma$ and activated α_i can open K^+ channels, although equivalent channel responses require a much higher concentration of $\beta\gamma$ than of activated α_i ^{12,15}. The Birnbaumer/Brown group strengthened the case for α_i by showing that K^+ channels are opened by recombinant α_i -protein made in bacteria¹⁶ and that a monoclonal anti- α_i antibody blocks G-protein-mediated opening of the channels¹⁷. It began to look as if champions of the α -chain were right, at least in contending^{1,16,17} that α_i mediates cholinergic regulation of cardiac K^+ channels.

Nonetheless, purified $\beta\gamma$ can open K^+ channels in patches of heart membrane^{2,15}. How does this come about? On page 557 of this issue⁷, Kim, Clapham, Neer and collaborators report evidence indicating that $\beta\gamma$ activates the channels by stimulating membrane-bound PLA₂. In patches of cardiac plasma membrane, an antibody previously shown to block PLA₂ activity prevents $\beta\gamma$ from opening K^+ channels; products of PLA₂ — arachidonic acid and arachidonic-acid metabolites (such as certain leukotrienes) derived by the 5-lipoxygenase pathways — also open the channels. Corroborating

G proteins, $\beta\gamma$ and phospholipase A_2

Cell type	Extracellular signal	Signal stimulates PLA_2	$\beta\gamma$ Stimulates PLA_2	Target of PLA_2 metabolites	Physiological importance	Refs
Neutrophil	Chemotactic peptide	Yes	Unknown	Unknown	Chemotaxis?	18
3T3 fibroblast	Thrombin	Yes	Unknown	Unknown	Mitogenic response?	19
Retinal rod cell	Light	Yes	Yes	Unknown	Unknown	13,21
Heart muscle	Unknown	Unknown	Yes	K^+ channels	Decreases heart rate?	7,8
<i>Aplysia</i> sensory neuron	FRMFamide	Yes	Unknown	K^+ channels	Synaptic inhibition	5,20
Haploid cells of <i>S. cerevisiae</i>	Mating pheromone	Unknown	Unknown	Unknown	Mating	3,4,6

evidence comes from experiments with a lipoygenase-blocking drug, nordihydroguaiaretic acid (NDGA); this drug prevents K^+ -channel opening in response to both $\beta\gamma$ and arachidonic acid. Kurachi *et al.*, on page 555 of this issue⁸, describe similar effects of arachidonic acid (also blocked by NDGA) and its metabolites on cardiac K^+ channels, but do not test for effects of $\beta\gamma$.

Ironically, the research group whose initial results² supported $\beta\gamma$ and seemed to exclude α_i now presents⁷ good pharmacological evidence that physiological regulation of K^+ channels by acetylcholine cannot be mediated by $\beta\gamma$: NDGA, the lipoygenase inhibitor that prevents opening of K^+ channels by $\beta\gamma$ and arachidonic acid, has no effect on channel opening triggered by cholinergic agonists. This negative result raises a puzzling question: does the effect of $\beta\gamma$ on channels in membrane patches relate to any physiological role of $\beta\gamma$ in the heart? One could imagine that the effects of $\beta\gamma$ and arachidonic acid serve as a backup signalling mechanism that may play a quantitatively greater part in tissues other than heart muscle; alternatively, such effects might reflect a signalling pathway triggered by some agent other than acetylcholine and mediated, perhaps, by a different G protein.

In either case — barring experimental artefact — the findings of Kim *et al.*⁷ suggest the existence of a novel G-protein-dependent signalling pathway, mediated by $\beta\gamma$, PLA_2 and lipoygenase metabolites of arachidonic acid. Although unproven, the existence of such a pathway brings together many previously unconnected observations (see table). Thus, G proteins mediate stimulation of PLA_2 in many cell types, including neutrophils¹⁸, fibroblasts¹⁹ and sensory neurons of the sea slug *Aplysia californica*^{5,20}; in each case, the signal is blocked by pertussis toxin, a bacterial ADP-ribosyltransferase that prevents G_i and transducin^{9,10} from interacting with their respective receptors.

Two cellular responses are especially attractive candidates for mediation by the postulated $\beta\gamma$ - PLA_2 -lipoygenase pathway (see table). In one of these^{5,20}, PLA_2 and its products are already known to be involved in signalling. The neuropeptide

Phe-Met-Arg-Phe-NH₂ (FRMFamide) inhibits synaptic transmission in an *Aplysia* sensory neuron by opening K^+ channels, thereby inducing the cell to hyperpolarize. Although the FRMFamide signal is known to be carried by a G-like G protein and to be transduced into opening of K^+ channels by lipoygenase metabolites of arachidonic acid, it is not clear which subunit of the G protein stimulates PLA_2 . The effects of $\beta\gamma$ on PLA_2 ^{7,13}, and of arachidonic-acid metabolites on cardiac K^+ channels^{7,8}, raise the possibility that $\beta\gamma$ mediates the action of FRMFamide in *Aplysia*, even though the cardiac and *Aplysia* K^+ channels differ pharmacologically and electrophysiologically.

Finally, it is tempting to suggest that mating pheromones may use a

$\beta\gamma$ - PLA_2 -lipoygenase pathway in *S. cerevisiae*, mainly because this is the only system (so far) in which the $\beta\gamma$ -subunit of a G protein has been clearly demonstrated⁶ to carry a hormonal message from the receptor to an intracellular effector pathway. No one knows whether the yeast effector is PLA_2 . Indeed, it is not clear (to me at least) that PLA_2 even exists in *S. cerevisiae*. Someone should take a look. □

1. Codina, J., Yatani, A., Grenet, D., Brown, A.M. & Birnbaumer, L. *Science* **236**, 442–445 (1987).
2. Logothetis, D.E., Kurachi, Y., Galper, J., Neer, E.J. & Clapham, D.E. *Nature* **325**, 321–326 (1987).
3. Dietzel, C. & Kurjan, J. *Cell* **50**, 1001–1010 (1987).
4. Miyajima, I. *et al.* *Cell* **50**, 1011–1019 (1987).
5. Volterra, A. & Siegelbaum, S.A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7810–7814 (1988).
6. Whiteway, M. *et al.* *Cell* **56**, 467–477 (1989).
7. Kim, D. *et al.* *Nature* **337**, 557–560 (1989).
8. Kurachi, Y. *et al.* *Nature* **337**, 555–557 (1989).
9. Gilman, A.G. *Rev. Biochem.* **56**, 615–649 (1987).
10. Birnbaumer, L. *et al.* *Kidney Int. (Suppl.)* **32**, S14 (1987).
11. Bourne, H.R. *Nature* **325**, 296–297 (1987).
12. Neer, E.J. & Clapham, D.E. *Nature* **333**, 129–134 (1988).
13. Jelsema, C.L. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2623–2627 (1987).
14. Herskowitz, I. & Marsh, L. *Cell* **50**, 995–996 (1987).
15. Logothetis, D.E., Kim, D., Northup, J.K., Neer, E.J. & Clapham, D.E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5814–5818 (1988).
16. Yatani, A. *et al.* *Nature* **336**, 680–682 (1988).
17. Yatani, A. *et al.* *Science* **241**, 828–832 (1988).
18. Okajima, F. & Ui, M. *J. biol. Chem.* **259**, 13863 (1984).
19. Murayama, T. & Ui, M. *J. biol. Chem.* **260**, 7226 (1985).
20. Piomelli, D. *et al.* *Nature* **328**, 38–43 (1987).
21. Jelsema, C.L. *J. biol. Chem.* **262**, 163–168 (1987).

Henry R. Bourne is in the departments of Pharmacology and of Medicine, University of California School of Medicine, San Francisco, California 94143–0450, USA.

Astronomy

Enlightenment and confusion

David Lindley

AFTER SEVERAL consecutive years in which the Space Telescope Science Institute in Baltimore had held a 'two years before launch' party, the resident astronomers were finally able, last December, to celebrate a scheduled launch only one year hence. But the return to active duty of the space shuttle should provide astronomers with much more than the long-awaited Hubble Space Telescope. Addressing a recent meeting*, Lennard Fisk, Associate Administrator for Space Science and Applications at NASA, promised a revitalized US space-science programme, covering astronomy at almost every part of the electromagnetic spectrum, with efforts in solar and planetary physics too.

Throughout the conference the notion that data from the Hubble Space Telescope (HST) or elsewhere would resolve theoretical conflicts or point in new directions was constantly heard. After years of relative drought, astronomy faces an imminent downpour of new information, to the extent that Fisk was concerned that there would be too few young scientists coming through the system to cope with

results from all the planned and scheduled instruments. Most of those attending the meeting seemed to find this a novel and agreeable problem to worry about.

New instruments might finally bring into view planets around other stars (D. C. Black, Lunar and Planetary Institute, Houston). An orbiting infrared telescope searching for brown dwarfs, bodies of only 0.1 solar masses with temperatures of a few thousand kelvin, may confirm the growing suspicion that almost every star has something orbiting around it — either another star, a brown dwarf or a planetary system. Obtaining images of distant Jupiters may even be possible: in visible light, a star would outshine such a planet by a factor of 10^9 but at infrared wavelengths the contrast could fall to 10^4 , a noise-to-signal ratio which astronomers do not find entirely overwhelming.

Recent infrared observations provide direct evidence that as many as 60 per cent of young Sun-like stars are accompanied by protoplanetary material (K. Strom, Five College Astronomy Department, Amherst). Theory indicates that such materials should turn into something else very quickly. Optically thick gaseous disks

*Annual meeting of the American Astronomical Society, Boston, 8–12 January 1989.

around stars seem to evolve on a timescale as short as 10^6 years, and simulations suggest that as a protoplanet pulls itself together out of the accreting material, its gravitational influence creates a stable gap in the disk, cutting off its own supply of material and putting a stop to its growth (D. N. C. Lin, Lick Observatory). High-resolution HST observations of T Tauri stars, believed to be protoplanetary systems, might be able to test such simulations directly. Conceivably the orderliness of our own system, as expressed by the simple empirical law discovered in the eighteenth century by Bode for the orbital radii of the planets, could be a consequence of hydrodynamics working on the density profile of the original gaseous disk. In this picture, Neptune lies where the disk became tenuous and optically thin, and convective flows ceased.

A bigger and more complicated accretion process may occur at the centre of our Galaxy. Although emission from the Galactic Centre is believed to be fuelled by gas falling into a central black hole of perhaps 10^6 solar masses, it is not so easy to see how the gas gets there. A clue was presented by P. T. P. Ho (Harvard University). Radio observations at 1-centimetre wavelength using the Very Large Array traced a faint streamer of emission in a line of ammonia, indicating a ribbon of cool gas starting about 10 parsec from the Galactic Centre and reaching down to the 2-parsec circumnuclear shell. Moreover, the outer end of the streamer seems to be associated with a giant molecular cloud, and a supernova remnant lurks nearby. This all suggests that a fairly recent supernova explosion disturbed the molecular cloud and set a parcel of gas moving towards the central black hole.

Other clues to the nature of the Galactic Centre are provided by observations of the γ -radiation emitted when electrons and positrons collide and annihilate (C. MacCallum, Sandia National Laboratories). First observed in 1977, the characteristic 511-kiloelectron volt line was confirmed in 1979 but seemed to go quiet later that year. In 1981 and 1984 it was still absent. Last year GRIS, a γ -ray telescope, was flown by balloon to observe supernova 1987A, but also rediscovered annihilation radiation from the Galactic Centre.

MacCallum recalled a related bonus discovery from the Solar Maximum Mission satellite which, when it was not looking for γ -rays from supernova 1987A, registered faint 511 kiloelectron volt emission distributed across the galactic plane (G. Share, Naval Research Laboratory; reported at the Texas Symposium on Relativistic Astrophysics, Dallas, 11–16 December 1988). The distributed emission could be due to energetic photons from the Galactic Centre creating occasional positrons in the denser material of the

galactic plane. The episodic emission from the Galactic Centre itself is probably connected with how the black hole is fed, perhaps signalling the end of a streamer of gas which set off earlier than Ho's.

In keeping with astronomical tradition, the most puzzling observation reported was one of the simplest. The famous void in Boötes, repeatedly surveyed and confirmed to be empty of optically detectable galaxies in comparison with surrounding regions, turns out to have at least a few galaxies visible to IRAS, the venerable infrared astronomy satellite (G. S. Aldering, University of Michigan). The IRAS survey preferentially picks out dusty galaxies, often with active star formation, and thus provides a different selection of galaxies than an optical survey. Work by M. Davis and M. Strauss (University of California, Berkeley) indicates that IRAS galaxies follow more or less the same distribution as optical galaxies, with the exception that densely clustered regions do not show up on the IRAS survey. But where there are empty spaces in the Center for Astrophysics survey, for example, there are similar deficiencies in the IRAS survey.

But in the direction of the Boötes void, much bigger and more distant than anything in the Center for Astrophysics survey, IRAS finds several galaxies, and

some of these, according to subsequent redshift determinations by R. Kirshner (Harvard University), are indeed in the void itself. So far only seven such galaxies have been found, but in comparison to the general distribution of IRAS galaxies this is not an unexpectedly low number for the volume of the void, meaning that what is certainly a void for magnitude-limited optical surveys is not a void for the IRAS survey. The significance of this is quite unknown. Modern cosmologists will not be too uncomfortable with the idea that two different surveys of galaxies will trace the distribution of mass in the Universe in somewhat different ways. What will not please them is the idea that IRAS and optical galaxies trace the mass similarly in one part of the sky but differently elsewhere.

Resolution of puzzles like this depends not only on the Hubble Space Telescope, but on other satellites and on bigger ground-based telescope. Observational cosmology becomes more puzzling the further the Universe is surveyed. As always, astronomers believe that enlightenment will come with the next round of instrumentation, but traditionally this enlightenment comes at the cost of another qualitatively new puzzlement. □

David Lindley is an assistant editor of Nature.

Ocean Drilling Program

The birth of the Indian Ocean

*Leg 123 shipboard scientific party**

THE early tectonic and palaeo-oceanographic history of the oceans is poorly understood. For this reason, we drilled two sites (765 and 766) off northwestern Australia (see map) with the following objectives: to elucidate the palaeo-oceanography, sedimentology and magmatic processes related to the rifting of the early Indian Ocean; to constrain the rift-to-drift tectonic history of one of the oldest oceanic basins; to refine the late Jurassic to early Cretaceous geological timescale; to establish

a geochemical reference site for oceanic sediment and basement composition in oceanic crust near a subduction zone.

In the early Mesozoic era (about 250 million years (Myr) ago), the western Australian margin was part of a continental rift zone on northeastern Gondwanaland that united Australia, Greater India and Antarctica. To the north of Gondwanaland lay the Tethys Sea. The geological transition from the Tethys into the modern Indian Ocean supposedly commenced with late Jurassic (perhaps 150 Myr ago) rifting and sea-floor spreading between north-west Australia and an unknown conjugate plate, which led to the formation of the Argo Abyssal Plain. This was followed in the early Cretaceous by rifting and separation of Greater India, western Australia and Antarctica, leading to connection of the early Indian Ocean with the early South Atlantic Ocean.

Rifting between Antarctica and Australia was initiated in the Cenozoic (approximately 65 Myr ago) and resulted in a steady northward drift of Australia, such that the present position of the north-western margin is near 16° S. In the mid-Tertiary (around 30 Myr ago), Australia

* Co-Chief Scientists: Felix Gradstein (Atlantic Geoscience Centre, Nova Scotia) and John Ludden (Univ. Montreal). ODP Staff Scientist: Andrew C. Adamson (ODP, Texas A&M Univ.). Also: Peter Baumgartner (Inst. Géologie, Lausanne), Roland Beaussillon (Centre de Recherches Géophysiques, Garchy), Tom Bolmer (Woods Hole Oceanographic Inst.), Paul Brown (University College London), Robin Brereton (British Geological Survey), Dick Buffler (Univ. Texas, Austin), David Castillo (Stanford Univ.), John Compton (Univ. South Florida), Julie Dumoulin (Univ. Calif., Santa Cruz), Cedric Griffiths (Univ. Trondheim), David Haig (Univ. Western Australia), David Heggie (Bureau Mineral Resources, Canberra), Akira Ishiwatari (Kanazawa Univ.), Michael Kaminski (Dalhousie Univ.), Kazuto Kodama (Kochi Univ.), David Kopaska-Merkel (Northeastern Science Foundation), Jean Philippe Marcoux (Univ. Paris), Andrew McMinn (New South Wales Geological Survey), Michael Moran (Univ. Nebraska, Lincoln), Jorg Mutterlose (Univ. Hannover), Brennan O'Neill (US Geological Survey), James Ogg (Purdue Univ.), Terry Plank (Lamont-Doherty Geological Obs.), Michael Riggins (Colorado School of Mines), Michael Schott (Univ. München), Gregory Simmons (Texas A&M Univ.) and Jürgen Thürow (Univ. Tübingen).

and the Indonesian Archipelago collided to form the New Guinea cordillera and the active subduction zone and associated ocean-floor trenches south of Indonesia.

We find that the Argo Abyssal Plain may be younger by 20 Myr than previously supposed, the plain opening up during the late Jurassic to early Cretaceous (163–125 Myr ago). Detailed microfossil stratigraphy for the brown-red silty claystone and volcanic ash layers, lying on fresh volcanic flows at site 765, dates the basement, previously assigned magnetically to the Oxfordian (about 156 Myr ago), as latest Berriasian–Valanginian (around 140 Myr). Equally, stratigraphic correlations with Deep Sea Drilling Project site 261, 250 km to the north suggest that the M23 (Kimmeridgian) basement at site 261 is oceanic crust of early Cretaceous age. These revised dates have important consequences for our understanding of the early evolution of the Indian Ocean.

Stratigraphic determinations based on microfossils and geomagnetic polarity in the basal sediments at site 766 indicate that sea-floor spreading between Australia and Greater India started at M11 time, 134 Myr ago, in the late Valanginian. The presence of a mid-ocean ridge basalt (MORB) type of tholeiitic intrusive sequence at the base of a sandy sedimentary section implies that the ocean/continent boundary must have been rifted, buried by continental sediments and intruded by MORB-type magmas immediately before the formation of true

oceanic crust a few kilometres seaward.

The calculated sedimentation was relatively high ($1.5\text{--}4\text{ cm kyr}^{-1}$) during the early oceanic-subsidence phase (Valanginian to Aptian; 138–113 Myr ago) at sites 765 and 766, but decreased by a factor of 2 in the post-Aptian. At site 766 this reflects the change from deposition of terrigenous material at river deltas to open-ocean blanket deposition (see figure). From the microfossil composition of the sediments it is clear that site 765 was further north than

previously estimated: 30° S rather than 45° S . This agrees with other estimates of the position of East Timor (Wensink, H. *et al. Geologie Mijnb.* 66, 89–101; 1987).

A major hiatus occurred at the end of the early Eocene at sites 765 and 766, which probably signifies the onset of vigorous deep-water circulation, owing to the formation of Antarctic deep water that eroded the abyssal sea floor. In mid-Tertiary time, several poorly understood factors combined to increase rapidly abyssal sedimentation in the Argo region

by a factor of 2–6. We postulate that canyons, like Swan Canyon near site 765, that align with the main south-west-north-east fault trend on the margin, became active sediment conduits at that time. This led to catastrophic Neogene abyssal sedimentation with massive debris flows, representing in part the tail of a large prograding clastic fan on the Exmouth Plateau (Apthorpe, M. in *The North West Shelf, Australia* (eds Purcell, P.G. & Purcell, R.R.) 55–84 (Petr. Soc. Austr., Perth, 1988)).

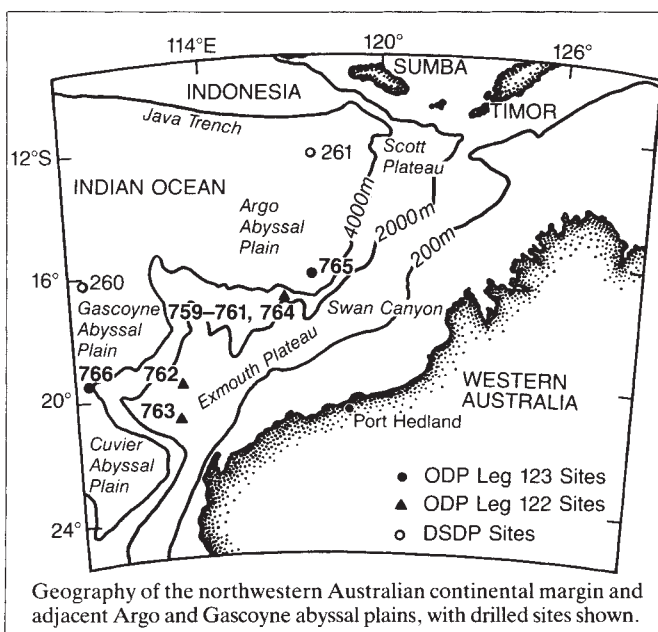
Few oceanic sites, and none in the Southern Hemisphere, exist that directly relate the M-series of geomagnetic reversals in the sediments with multiple biostratigraphy. Sites 765 and 766 yielded a detailed

magnetic polarity sequence for M13–M0, Valanginian to Aptian which, together with the detailed microfossil record and potential radiometric dates of fresh basalt underneath and volcanic ash in the basal sediments, will enhance the Mesozoic chronology.

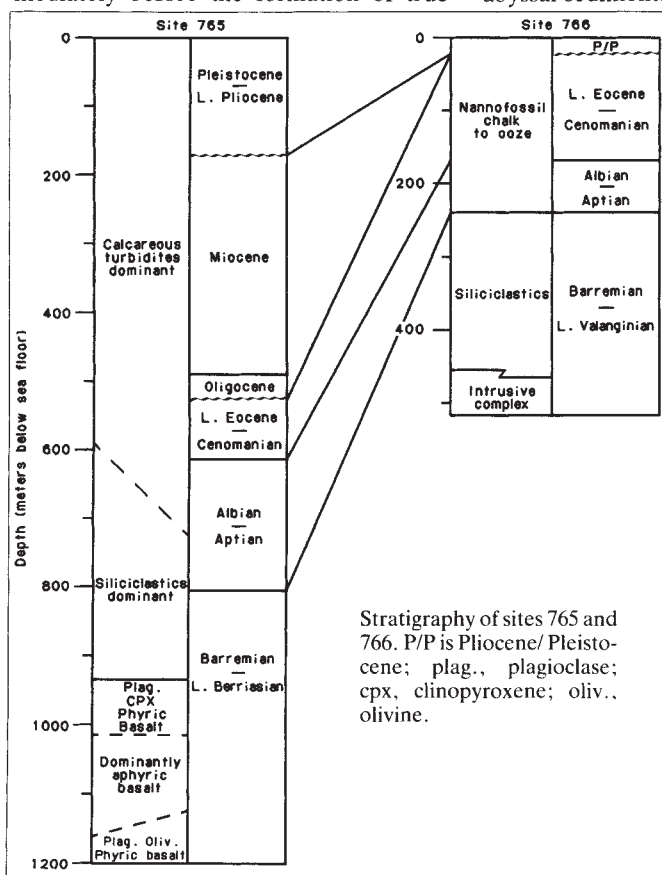
The first of a series of geochemical reference sites was drilled at site 765. Shipboard X-ray fluorescence analyses of a comprehensive suite of bulk sediment and clay fractions were complemented by geochemical logs for the elements K, U, Th, Ca, Si, Fe and Al for an entire hole. Most of the variation is a result of simple mixing between three end members: biogenic calcium carbonate, biogenic silica and an aluminosilicate component. Geochemical studies of the terrigenous and volcanogenic clays will be used to fingerprint changes in the bulk composition of the oceanic crust.

The basalts erupted at the Indian Ocean Ridge axes have subtle geochemical differences from those in the Pacific and Atlantic oceans so that each region must have a distinct mantle source. One of our main petrological objectives was to determine how Indian Ocean basalt chemistry has changed during the past 150 Myr. Despite their early Cretaceous age, site 765 lavas have preserved fresh glass, making shore-based isotope evaluation of the earliest Indian Ocean mantle reservoirs possible. Shipboard geochemical data for these basalts indicate significant differences in incompatible element geochemistry between these basalts and those erupted in the modern mid-Indian Ocean Ridge axis.

This leg is the last of nine successful Indian Ocean drilling expeditions of the ODP. Site 765 contains the deepest cased oceanic hole in the world and, with its re-entry cone on the sea floor, it is in perfect shape for future drilling. □



Geography of the northwestern Australian continental margin and adjacent Argo and Gascoyne abyssal plains, with drilled sites shown.



Stratigraphy of sites 765 and 766. P/P is Pliocene/Pleistocene; plag., plagioclase; cpx, clinopyroxene; oliv., olivine.

Copulation dynamics

Out for the sperm count

Paul H. Harvey and Robert M. May

WHY does testes size vary among species? In a series of seminal papers¹⁻³, Roger Short developed the idea that differences in testes size among the great apes result from sperm competition. A chimpanzee is one-quarter the weight of a gorilla, yet its testes are four times heavier. Short noted that female chimpanzees, unlike female gorillas, often mate with more than one male during a single oestrus, resulting in a selective premium on males to be successful at sperm competition. One way to increase the probability of being successful at sperm competition is to produce more sperm, hence the chimpanzees' larger testes. Subsequent comparative studies demonstrated that Short's idea is not restricted to the great apes. Across primates as a whole, those species in which females live in multi-male societies are the ones in which males have large testes for their body size^{4,5}. Casting the taxonomic net more widely, similar patterns have now been found in other mammals^{6,7} and in birds⁸.

How is testes size related to rates of sperm production and to sperm proficiency? By analysing his data in such a way as to hold the effects of body size constant, Anders Møller finds⁹ that primates living in multi-male societies not only have larger testes, but that they also produce larger volumes of ejaculate containing more sperm. Furthermore, the proportion of those sperm that are motile is higher among the multi-male species, varying from around 90 per cent in some multi-male macaques to a meagre 7 per cent in the monogamous gibbons. This finding apparently argues against the suggestion that non-motile sperm are 'kamikaze' sperm which act merely to facilitate the development of copulation plugs¹⁰. If that were so there should be higher proportions of non-motile sperm in multi-male

species, where the production of copulation plugs is more worthwhile.

It seems to us that the presence of sperm competition provides an alternative to the usual explanation for sexual swellings which advertise oestrus in female primates that typically live in multi-male groups. It could be that, rather than provoking physical combat among males or enabling females to choose among prospective mates¹¹, such swellings promote copulations by more than a single male. If there is genetic variance for success at sperm competition, there will be selection for females to choose males whose sons are likely to be successful at sperm competition. One way to achieve that end is directly to test putative fathers against each other.

In his new work on primates, Møller considered rates of sperm transmission. He investigated sperm production by locating data on rates of production for nine species of mammals, including humans¹². Among those species, neither sperm production rate per gram of sperm-producing tissue nor sperm transit time through the epididymis increases with testes size. But after correction is made for body size, species with larger testes have higher sperm-production rates, larger sperm reserves, and they release more sperm per ejaculate. The number of ejaculate equivalents held in sperm reserves (that is, sperm reserves/sperm per ejaculate) seems unrelated to body or testes size. On average, reserves hold sperm equivalent to about 21 ejaculates but with very large differences among species. For example, sheep hold reserves equivalent to about 95 ejaculates, rabbits about 30, but humans less than 2. Presumably such differences are associated with the likelihood of males getting the opportunity to remate in a short time period and

with the potential for the mating to provide the opportunity of producing additional offspring.

Human males, as well as having low sperm reserves, also have a low rate of sperm production per gram of sperm-producing tissue. Møller found¹² that the daily sperm production rate per gram of parenchyma varied from 12×10^9 to 25×10^9 across the eight non-human mammals, whereas humans produce a mere 4.25×10^9 sperm per gram. Most studies on humans have been performed on Caucasians, but there are marked differences in testes size among human races. Even controlling for age differences among samples, adult male Danes have testes that are more than twice the size of their Chinese equivalents^{13,14}, for example. This difference is much greater than would be expected on the basis of racial differences in body size. Various estimates¹³ suggest that individual Caucasians produce about twice the number of spermatozoa per day than do Chinese ($185-253 \times 10^6$ compared with 84×10^6).

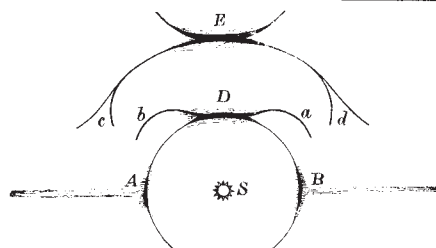
Despite this difference and the evidence from other primates, there is a reluctance to connect differences in human testes size with historic differences in mating patterns and their associated differences in degrees of sperm competition^{13,14}. Primates living in single-male groups, for example, have relatively small testes of similar size, irrespective of the number of females in the group (so-called 'monogamous' species have testes the same size as 'harem' species). There is no evidence relating testes size to sperm competition among human races¹⁴, perhaps because nobody has looked for it. Rough estimates of uncertainty in paternity are available for different human cultures, but it is not known whether they relate to testes size when body-weight effects are controlled for. We think it is as likely that testes size differences among human races have been adaptive in their own right — different responses to different mating behaviours — as that the differences arose as a correlated non-adaptive response for selection of ovary size and twinning rates among females¹⁵. In neither case, however, has a statistically sound comparative test been performed.

The evidence from bird species seems more equivocal. After the usual correction for body size, there is a positive correlation between testes size and polygyny^{8,15}. But this does not bear directly on the issue of whether testes size is correlated with sperm competition; polygyny by no means necessarily implies sperm competition, nor does monogamy imply its absence. Recent use of phenotypic correlations and molecular techniques, such as electrophoretically detectable protein variation and DNA fingerprinting, show evidence for high levels of adultery among some monogamous species. Westneat's study¹⁶

100 years ago

BETWEEN 1 and 2 p.m. of January 11, a solar halo, so remarkable as to deserve some notice in the columns of *Nature*, was observed and sketched by myself and several of my pupils. The mock suns A, B, and D (see diagram), appeared to be at the usual distance of about $22\frac{1}{2}^\circ$ from S, and the halo at E about the same distance from D. A and B were quite bright, but D and E were nearly twice as brilliant, and blazed with gorgeous prismatic colours.

The parhelic circle — observed by Prof. William Ellis on April 1, 1886 (*Nature*, vol. xxxiii. p. 535) — was very bright. It extended only from the mock suns A and B outwards from S to about 120° from the latter; and on the right branch of this circle was another mock sun (not shown in diagram) at the distance of about 90° from B. This last sun, as well as the visible



portions of the parhelic circle, was formed of pure white light, and the latter was everywhere parallel with the horizon. But perhaps the most remarkable part of the phenomenon was the forking of the arc *c E d* at the ends *c* and *d*, and the concave recurving of the arc *a D b* (convex to S at D) at the ends *a* and *b*. EVAN McLENNAN.

From *Nature* 39, 341; 7 February 1889.

of the predominantly monogamous indigo bunting *Passerina cyanea* reveals that at least 14 per cent of 256 offspring sampled have genotypes incompatible with one of the putative parents. Because females do not lay eggs in each other's nests in this species, extra-pair copulations are probably responsible for fathers helping to raise offspring that are not their own (although in the indigo bunting, direct paternal care is minimal).

Such findings have tended to provoke a revisionist movement, replacing the view that 90 per cent of bird species are monogamous with the view that essentially none are. But if a male has a sufficiently high probability of fathering the offspring in his territory, selection will favour him helping to raise them. Some monogamous species copulate more frequently than others, with the frequently copulating species having relatively larger testes but fewer sperm per ejaculate⁸. Perhaps monogamous birds can be divided into two categories: those with few extra pair copulations and reduced sperm competition (infrequent maters); and those with relatively high levels of extra-pair copulations and sperm competition (frequent maters).

Because the last male to mate is more likely to fertilize the eggs, the male's strategy in those species with high levels of extra-pair copulations would be to mate with his mate frequently, and particularly after she has taken part in an extra-pair copulation, as in the case with indigo buntings¹⁷ and zebra finches¹⁸. Because, in zebra finches at least, the sperm of the last male to copulate have a precedence of about 84 per cent, and pairs mate about 12 times during the production of a clutch, such species may be as close to monogamy as can be reasonably expected in an imperfect world. □

- Short, R. V. In *Proceedings of the Canberra Symposium on Reproduction and Evolution* 3-19 (Australian Academy of Sciences, 1977).
- Short, R. V. *Adv. St. Behav.* **9**, 131-158 (1979).
- Short, R. V. in *Reproductive Biology of the Great Apes* (ed. Graham, C. E.) 319-341 (Academic, New York, 1981).
- Harcourt, A. H. *et al. Nature* **293**, 55-57 (1981).
- Harvey, P. H. & Harcourt, A. H. in *Sperm Competition and the Evolution of Animal Mating Systems* (ed. Smith, R. L.) 589-600 (Academic, New York, 1984).
- Clutton-Brock, T. H., Guinness, F. E. & Albon, S. D. *Red Deer: Behaviour and Ecology of Two Sexes* (Chicago University Press, 1982).
- Kenagy, G. J. & Tromlak, S. C. *J. Mammal.* **67**, 1-22 (1986).
- Møller, A. P. *Biol. J. Linn. Soc.* **33**, 273-283 (1988).
- Møller, A. P. *J. hum. Evol.* **17**, 479-488 (1988).
- Baker, R. R. & Bellis, M. A. *Anim. Behav.* **36**, 936-939 (1988).
- Clutton-Brock, T. H. & Harvey, P. H. in *Growing Points in Ethology* (eds Bateson, P. P. G. & Hinde, R. A.) 195-237 (Cambridge University Press, 1976).
- Møller, A. P. *Funct. Ecol.* **3**, 91-96 (1989).
- Short, R. V. in *One Medicine* (ed. Ryder, O.) 32-44 (Springer, New York, 1984).
- Diamond, J. M. *Nature* **320**, 488-489 (1986).
- Cartar, R. V. *Naturwissenschaften* **72**, 157-158 (1985).
- Wesneat, D. F. *Anim. Behav.* **35**, 865-876 (1987).
- Wesneat, D. F. *Anim. Behav.* **35**, 877-886 (1987).
- Birkhead, T. R., Pellatt, J. & Hunter, F. M. *Nature* **334**, 60-62 (1988).

Paul H. Harvey and Robert M. May are in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Nikolaas Tinbergen (1907-1988)

NIKOLAAS Tinbergen, who received a Nobel prize for Biology and Medicine in 1973, died on 21 December 1988. As a student in the Netherlands Tinbergen was an outstanding hockey player, but was not otherwise regarded as showing great promise. He nonetheless began studies on several species of birds and insects, in which he quickly revealed an exceptional capacity for meticulous, detailed observation. An autobiographical volume, *Curious Naturalists* (1958), gives an agreeable account of many of the field enquiries in which he took part. At the end he typically questions the value of what he had been doing, but concludes that nobody "need be ashamed of being curious about nature".

Tinbergen, like Darwin, was interested not only in the variety of living things, but also in developing general concepts which allow us to impose order on diversity, and especially in proposing how the many forms of species-typical behaviour have arisen. But, unlike Darwin, he was also a skilful experimenter: his elegant experiments on insects and birds in the field, and on sticklebacks in aquaria, remain models for textbooks of ethology.

His large output of published work, part of which is collected in the two volumes of *The Animal in its World* (1972, 1973), includes many papers on the analysis of fixed action or motor patterns, especially social signals. These are accompanied by scrupulous comments on method, and provide much of the foundation of current social ethology. Empirical findings on behaviour are presented in a darwinian framework; but Tinbergen distinguished, with exemplary clarity, between experimentally validated findings, on the one hand, and hypotheses about the past action of natural selection, on the other.

Tinbergen's studies of gulls, carried out with many colleagues, illustrate the several facets of his method. Long periods of watching and filming in the field led to intimate understanding of the natural history of each species. Ingeniously designed models were used to analyse social interactions, including those of parents and young, and to reveal the significance of minutiae, such as the coloured spot on a herring gull's beak. An ecological component is exemplified by investigations on defence against predators. Among the items closely scrutinized was the removal of broken eggshells from the nest. This apparently trivial manoeuvre, reported on in three papers, led to discussion of difficult questions on the nature and functions of stereotyped acts, and of the complexity of the 'anti-predator system' of black-headed gulls.

Although, as Tinbergen wrote, even "common descent is a matter of conjecture", one aim was always to suggest how the behaviour recorded had evolved. Hypotheses on phylogeny were made plausible

by comparing the behaviour patterns of a number of closely related species.

Tinbergen moved from Leiden to Oxford in 1949, and there founded a school of able pupils among whom he inspired affection as well as respect. Their influence is now felt throughout the world. Two years later *The Study of Instinct* was published; this, the short Methuen monograph *Social Behaviour in Animals* (1953), and his notable study of a single species, *The Herring Gull's World* (also 1953), provided the themes for many new university courses in animal behaviour. Tinbergen was, indeed, a teacher of the front rank: as he wrote, he was "at heart a missionary".

Converting the English to ethology was easy. American behavioural scientists, accustomed to laboratory experiments on habit formation by domestic species, found the untidy world of nature disconcerting and the study of 'instinct' baffling. They were therefore slower to accept Tinbergen's message, but eventually did so on a large scale. Tinbergen's success in the United States was probably helped by his lecturing style, for he held close attention without flamboyance or gimmickry: the art that conceals art. Some appreciative North Americans also offered cogent criticisms of early ethological theory, especially the neglect of ontogeny. These Tinbergen discussed and partly accepted without strident controversy — an aspect of his readiness to modify his views as knowledge advanced. As a result, leading North American experimental psychologists joined the International Ethological Congress, of which Tinbergen was a prominent founder.

Tinbergen's characteristic restraint was also shown in his contributions to the human sciences. He held that a combination of modern ethology with darwinism could tell us much of value about humanity, from the social lives of infants to the anti-social violence of adults; and he was tolerant of "naked-aper" among his colleagues and pupils. But he did not take part in the movement to provide a thoroughgoing biological interpretation of humanity. In 1968 he wrote: "What we ethologists do not want, what we consider definitely wrong, is uncritical application of our results to man".

Today, it may seem strange that a career in behavioural science of historical significance could be founded entirely on observation of whole animals, with little physiological analysis and no use of the algebra of evolutionary theory. A fitting memorial would be for students and workers in the behavioural sciences to return to Tinbergen's works, and to study them with the kind of critical appreciation which he himself applied to the works of others.

S.A. Barnett

S.A. Barnett is at the Australian National University, PO Box 4, Canberra, ACT 2600, Australia.

Pulsar astronomy

Spin-up more apparent than real

Andrew Lyne

GLOBULAR clusters of stars have recently been found to be a rich source of millisecond pulsars¹⁻³. In a search for such objects, Wolszczan *et al.*, who report on page 531 of this issue⁴, have discovered a pulsar, PSR2127+11, near the centre of the globular cluster M15, one of the most dense clusters in the northern skies. What makes this pulsar interesting is that, not only does it have a relatively long spin period, but it is solitary and, unlike any other known radio pulsar, its rotation rate seems to be increasing. The authors conjecture that this apparent spin-up may be an illusion due to the acceleration towards us of the object in the gravitational field of the collapsed core of the cluster. There seem to be two mysteries which need to be explained: why is the acceleration so large and why is the pulsar solitary?

Globular clusters are gravitationally bound swarms of 10^5 – 10^6 old stars found in our Galaxy. The central stellar density is often as high as 10^4 – 10^5 stars per cubic parsec compared with less than 1 pc^{-3} in the solar neighbourhood. This relatively large density probably causes many stellar interactions, resulting in a high incidence of low-mass X-ray binary systems in which we see compact objects, probably neutron stars, being spun up by accretion of material from less evolved companions. We detect many of these systems by the copious X-rays emitted by the hot accreting gas. This is the process by which it is believed old long-period radio pulsars are given the rapid rotation rates which we subsequently see in millisecond pulsars after the accretion has ceased (see ref. 5). Hence the success of the searches for such pulsars in globular clusters carried out recently.

The new pulsar has a period P of 110 ms so that it is not a true millisecond pulsar and the absence of varying Doppler effects indicates that it is not in a short-period binary system. But although all other known pulsars spin down as they lose rotational kinetic energy through their radiation, PSR2127+11 is observed to be spinning up slightly. This cannot be due to any accretion process as it has no companion and so it seems probable that the observed spin-up ($\dot{P} = -2 \times 10^{-17} \text{ s s}^{-1}$) is only apparent. It is likely to be the result of the acceleration of the pulsar towards the Earth in a gravitational field, either that of its nearest neighbour or that of the collapsed core of the cluster: the chain of regular pulses from the object becomes increasingly compressed from our point of view. The implied acceleration of about

$5 \times 10^{-8} \text{ m s}^{-2}$ is, on the other hand, rather large for either of these explanations; the pulsar needs to be improbably close to its nearest neighbour or the core must be substantially more massive than we believe. Precise timing measurements over the next few years and the determination of the second period derivative could allow the distinction to be made between these two rather uncomfortable possibilities.

Although not a millisecond pulsar, PSR2127+11, with a period of 110 ms, is spinning rapidly compared with normal old pulsars. Together with the seemingly small value of intrinsic spin down rate, this suggests that the pulsar is an old one which was rejuvenated somewhat by accretion in a binary system. This leaves the question as to what has happened to the donor companion star. Wolszczan *et al.*⁴ suggest two possible explanations. First, the companion star has been ablated by radiation from the pulsar, a mechanism proposed⁶ to explain the recently discovered eclipsing binary pulsar PSR1957+20. However, they argue that it is unlikely that this pulsar was ever spinning rapidly enough to have the required energy output to achieve this. Second, perhaps the binary system was disrupted by an encounter with a third star⁷.

There is a third possibility that should be considered: the disruption of the binary system by the supernova explosion of the companion. The circumstances of the

explosion would be similar to those involved in the formation of the celebrated Hulse–Taylor system containing the binary pulsar PSR1913+16 and another neutron star⁸. In this case, the binary system survived the supernova explosion in which the first neutron star was formed. The companion subsequently evolved, and overflowed its Roche lobe (beyond which its gravity fails to hold back material), so spinning up the neutron star, before it underwent a supernova explosion itself. In the case of PSR1913+16, the binary system was not quite disrupted and the old pulsar with a period of 59 ms was left in a very eccentric orbit around the new neutron star formed by the exploded companion.

The solitary pulsar PSR2127+11 could have been formed in circumstances only slightly different: the second supernova could have disrupted the binary system and, provided it was not ejected from the cluster, PSR2127+11 would have been left in its intermediate-period spin state. If this was the case, it must have happened a long time ago when the cluster was young and there were still young, massive stars which could suffer supernova collapse. □

1. Lyne, A.G. *et al.* *Nature* **328**, 399–401 (1987).
2. Lyne, A.G., Biggs, J.D., Brinklow, A., Ashworth, M. & McKenna, J. *Nature* **332**, 45–47 (1988).
3. Ables, J.G. *et al.* *IAU Circ.* No.4602 (1988).
4. Wolszczan, A. *et al.* *Nature* **337**, 531–533 (1989).
5. Van den Heuvel, E.P.J. in *The Origin and Evolution of Newton IAU Symposium* No.125 (eds Helford, D.J. & Huang, J.H.) 393–406 (Reidel, Dordrecht, 1987).
6. Fruchter, A.S., Stinebring, D.R. & Taylor, J.H. *Nature* **333**, 237–239 (1988).
7. Rappaport, S., Putney, A. & Verbunt, F. *Astrophys. J.* (submitted).
8. Hulse, R.A. & Taylor, J.H. *Astrophys. J.* **195**, L51 (1975).

Andrew Lyne is at the University of Manchester Nuffield Radio Astronomy Laboratories, Jodrell Bank, Macclesfield SK11 9DL, UK.

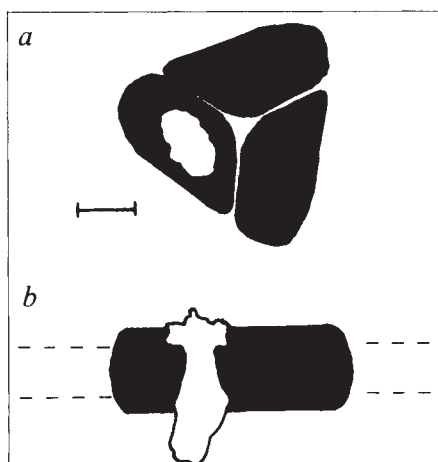
Photosynthesis

Giant revelations to come

Bob Ford

CONSIDER a huge, membrane-located protein complex that is so complex that about a quarter of its mass consists of tightly bound pigment molecules. Imagine this protein assembly also contains an intricate electron pathway which, using light energy, produces stable reduced components at redox potentials around minus 500 mV. This is no fictitious giant, but one of the two light-energy-utilizing systems of plant photosynthesis, the photosystem I reaction centre. Ingrid Witt and her collaborators have now succeeded¹ in showing that crystals of this reaction centre can diffract X-rays to 4 Å. Moreover, the hexagonal crystals are relatively stable in the X-ray beam, which will reduce the problems for data collection. It seems likely that the atomic structure of this big but beautiful protein will soon be obtained.

As a source for their reaction centre, Witt *et al.* use one of the most thermophilic photosynthetic organisms known, the vivid blue cyanobacterium *Synechococcus*, found in a Japanese hot spring at temperatures of around 70 °C. The reaction centre crystallizes as a trimeric complex which has a total relative molecular mass of about 800,000. Each of the three reaction centres contains about 60 chlorophyll *a* and 10 carotenoid molecules, giving the crystals an intense green colour. The overall form of the cyanobacterial complex is known from electron microscopy (see figure). Witt *et al.* present a model for the organization of the complexes within the crystals in which the repeating unit of dimensions 285 × 285 × 167 Å is composed of two pairs of trimeric complexes.



Simplified representations of the trimeric photosystem I reaction-centre complex isolated from cyanobacteria viewed from the top (a), and from the side (b). For size comparison, the outline of the purple bacterial reaction centre from *Rhodospseudomonas viridis* is shown in white. The approximate boundaries of the lipid bilayer membrane are marked with dashed lines. Bar, 5 nm.

What surprises does the structure of this reaction centre hold? It seems certain that the structure will be radically different from the structure of the purple bacterial reaction centre, the elucidation of which² led to last year's Nobel prize in chemistry. First, there is no similarity in the amino-acid sequence between any of the protein components of the purple bacterial and photosystem I reaction centres. Second, the final electron acceptors are completely different. Last, and most important, an integral part of the photosystem I reaction centre must be much larger than the purple bacterial reaction centre (see figure), and it seems likely that in the membrane-spanning regions of the two main protein subunits of the photosystem I reaction centre, almost every amino acid will be within striking distance of a non-covalently bound chlorophyll or carotenoid molecule. It may even turn out that these rather rigid pigment molecules contribute to the overall structure of the protein complex.

One question not answered by the work on the purple bacterial reaction centre is how light excitation energy is transferred from accessory pigments to the primary electron donor. This crucial question should now be answerable by resolving the spacing and orientation of the pigment molecules in the photosystem I reaction centre. A further interesting feature of the light-harvesting system is that a fraction of the chlorophylls actually absorb at longer wavelengths than the primary electron donor. Thus, if they are to function in light-harvesting, they must apparently transfer their energy 'uphill' to wavelengths a few nanometres shorter. It is already known from linear dichroism measurements that these long-wavelength-absorbing chlorophylls have a particular

orientation within the complex. Perhaps this intriguing problem will also be resolved by determining the structure at atomic resolution.

What will the new work reveal about photosynthesis in plants? The strongly reducing components of the reaction centre drive an electron-transfer chain which results in CO₂ fixation, and also provide power for N₂ fixation in cyanobacteria. Fortunately, there seems to have been very little variation in the photosystem I reaction centre throughout evolution³, and therefore the structure obtained from a cyanobacterial source should be widely applicable to plants. But there are inevitably some differences between the crystallized complex and the system existing in the native membranes. First, it is actually the monomeric form of the reaction centre which exists in the native cyanobacterial membranes^{4,5} rather than the trimeric complex which forms after extraction from membranes. It is not known why the trimer forms in these

conditions, but as it is the only form of the reaction-centre complex which crystallizes, this property is a remarkable gift for crystallographers trying to determine its function. Second, although the reaction centre is completely active in the primary reactions of light-induced electron transfer, it has lost to some extent its ability to communicate rapidly with the secondary (water-soluble) electron acceptors and donors. This is probably because of the loss during the isolation procedure of some of the protein subunits that are suspected to be involved in mediating these processes. □

1. Witt, I. *et al. Ber. Bunsenges. Phys. Chem.* **92**, 1503–1506 (1988).
2. Deisenhofer, J. *et al. Nature* **318**, 618–624 (1985).
3. Cantrell, A. & Bryant, D. A. *Plant. molec. Biol.* **9**, 453–468 (1987).
4. Ford, R. C. & Holzenburg, A. *EMBO J.* **7**, 2287–2293 (1988).
5. Mörschel, E. & Schatz, G. H. *Planta* **172**, 145–154 (1987).

Bob Ford is in the Department of Biophysical Chemistry, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.

Palaeontology

Fossil fishes from Gogo

Per Ahlberg

Fossil fishes from the Gogo formation of Western Australia are preserved in outstanding anatomical detail, providing unique opportunities for the study of evolution. John Long, for example, has now described a new genus of fish from Gogo¹ and, in another paper, reports the discovery of a fish² that should help to resolve a long-standing controversy about the evolution of land vertebrates.

Located in a remote part of Western Australia, the Gogo formation was first explored by Harry Toombs in 1963 (Fig.

1). The formation was deposited in the late Devonian period, about 360 million years ago, during a critical time in vertebrate evolution when all the main fish groups had emerged but land vertebrates (tetrapods) were yet to appear. The importance of the Gogo fossils lies in their excellent preservation.

Fishes are soft-bodied animals, and during fossilization the weight of the sediments tends to flatten the carcass. But the Gogo fishes are fossilized 'in the round'; all bones retain their natural

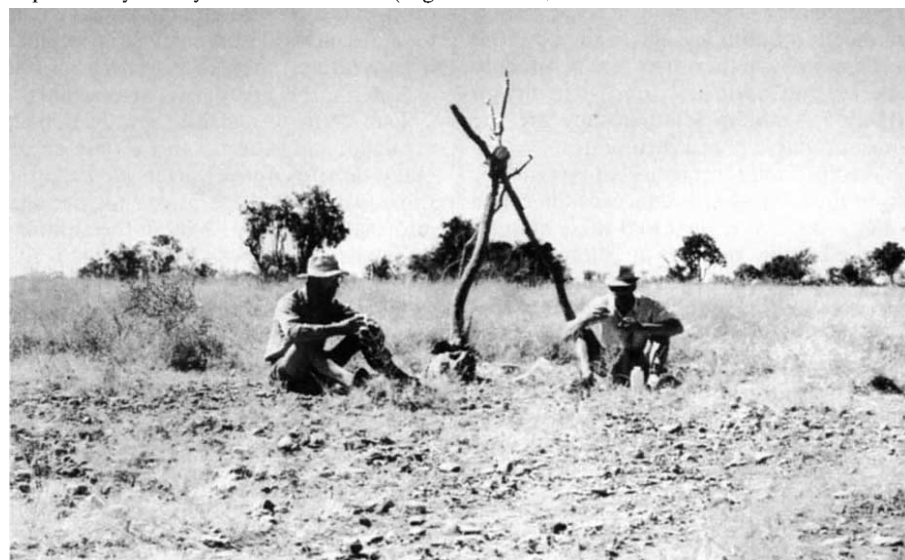


Fig. 1 The late Harry Toombs (left) unearthed many important specimens from the Gogo formation on behalf of the British Museum (Natural History) and pioneered acid-preparation techniques to free the fossils from their calcareous matrix. □



Fig. 2 *Latocamurus coulthardi*, a new species from Gogo reported by Long¹. Bar, 1 cm.

shape, and the spaces for the tissues are perfectly preserved inside the skull. As the rock is calcareous, it can be dissolved away from the bones with dilute acetic acid. This technique is rapid and delicate, yielding specimens which are comparable in quality with present-day skeletons. Three-dimensional fossil-fish specimens are often studied by serial grinding, where the fossil is 'sectioned' by grinding it away, and the 'sections' are used as templates for a large wax model of the specimen. This method yields a detailed picture, but destroys the fossil, and translation from sections to a wax model is a potential source of error. In an acid-prepared specimen from Gogo, on the other hand, the original anatomical structures are clearly visible, and the quality of preservation means that virtually no reconstruction or interpretation is needed. It is this degree of objectivity which makes the Gogo fossils invaluable.

In the first of his new papers¹, Long describes a new genus of placoderm, an extinct group of armoured fishes whose jaws were armed with bony shearing blades rather than normal teeth (Fig. 2). In the newly discovered form, the fragile anterior part of the braincase is perfectly preserved, so the exact relationship between this structure and the superficial bones can easily be reconstructed. The relationships of the placoderms are uncertain: opinions vary as to whether they are related to sharks and ratfishes, to osteichthyans (bony fishes), or whether they are the most primitive jawed vertebrates.

Another controversy fuelled by evidence from the Gogo fossils concerns the origin of tetrapods. It is clear that these animals evolved from one of the 'lobe-finned' bony fish groups, but at present there is a vigorous debate about the exact relationship. The traditional view links tetrapods with the extinct osteolepiform fishes, but Rosen *et al.* in 1981 claimed instead that lungfishes are the closest relatives of tetrapods³. Their reinterpretation focused on the choana, the internal nostril which in tetrapods opens from the nasal sac into the mouth. As this structure is absent in most fishes, including some of the lobe-fins, any fish with a choana must be related to tetrapods. Modern lungfishes have an internal nostril, and the grinding method

reveals that osteolepiforms have a canal leading from the nasal sac to the palate in the correct region⁴ — both groups could therefore be choanate. But the tetrapod choana is always surrounded by a particular set of bones present in osteolepiforms but absent in modern lungfishes, and this has been taken as evidence that the lungfish internal nostril is not a choana.

Some of the Gogo lungfishes have internal nostrils surrounded by bones⁵. Gardiner and Rosen independently realized that these bones could be the same as those of tetrapods, and that the internal nostril could, after all, be a true choana. This led to their reassessment³ of the evidence for a tetrapod-osteolepiform relationship, and their rejection of many of the characters supposedly uniting the two groups. Notably, they rejected the view that osteolepiform fishes were choanate — the size of the supposed choana seemed too variable, and the serial grinding method on which the reconstructions were based left too much room for error. Their conclusions have been fiercely debated⁶, despite the lack of new anatomical data since Rosen *et al.* published their conclusions. Long's second new paper² provides that evidence. He describes a new osteolepiform fish from Gogo. This fossil definitely has an opening on the palate large enough to have contained a nostril, supporting the original interpretation that the serially ground osteolepiforms are related to tetrapods.

The controversy has not yet been resolved, and in any case it is only one of many debates about vertebrate relationships. To solve these problems, detailed information is needed about the anatomy of fossil vertebrates, information which can be gained only from exceptionally well-preserved specimens. The Gogo fossil fishes provide such information; their scientific value is incalculable. □

1. Long, J. A. *Zool. J. Linn. Soc.* **94**, 233–258 (1988).
2. Long, J. A. *Natn. Geogr. Res.* **4**, 436–450 (1988).
3. Rosen, D. E., Forey, P. L., Gardiner, B. G. & Patterson, C. *Bull. Am. Mus. Nat. Hist.* **167**(4), 159–276 (1981).
4. Jarvik, E. *Zool. Bidrag Uppsala* **21**, 235–675 (1942).
5. Miles, R. S. *Zool. J. Linn. Soc.* **61**, 1–328 (1977).
6. Panchen, A. L. & Smithson, T. R. *Biol. Rev.* **62**, 341–438 (1987).

Per Ahlberg is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Daedalus

Real dry cleaning

WASHING is ridiculously inefficient. Grams of detergent and kilograms of hot water are energetically agitated in order to remove a few milligrams of dirt; and the whole mixture is then just thrown away.

Seeking a better process, Daedalus recalled those wonderfully repulsive children's toys made of a highly plasticized and very sticky rubber. When dropped on a typical kitchen floor, they adhere strongly and then slowly peel off, bearing a fine catch of crumbs, dust, fluff and hairs. But as dirt-removers, they only shift the problem along: the manufacturers advise you to wash them clean with soap and water.

So DREADCO's chemists are extending the principle. They are experimenting with derivatives and analogues of gelatine, plasticized with water into a highly elastic gel. The addition of a suitable viscous hydrophilic polymer for stickiness, and a detergent to bind dirt particles, gives a new consumer product: DREADCO's Dry Soap. A cylinder of Dry Soap mounted in paint-roller fashion will clean walls and floors and windows and cars and furniture with wonderful speed and efficiency. It will need a somewhat spiky or corrugated surface, both to improve its reach into small crevices and to enlarge its dirt-capacity; even so it will soon get covered with grime. So the DREADCO team are giving it a low melting point. A used Dry Soap roller can simply be melted, and poured back into its corrugated-cylinder mould. The dirt diffuses into the bulk of the gel, and the cleaning surface is restored. When after much use the whole volume of the roller is saturated with dirt, a melt-filtering step regenerates a new roller, and recovers all the dirt. Nothing is wasted!

This elegant new technology saves water, reduces detergent effluent, and simplifies the cleaning process. Housewives and domestic cleaners everywhere should throw down their mops and brooms and buckets to welcome it with open arms. On the larger scale, big Dry Soap rollers hauled by tractors could clear grime and litter from streets and parks; while tiny ones might make ideal erasers for ink and ball-point writing. Dry Soap may even be suitable for cleaning clothes and fabrics. Of course, even the most convoluted sticky roller is not going to penetrate far into woven material. But dirt so deeply entrapped that it cannot be reached from the surface is presumably harmless anyway.

The personal-hygiene market is less promising. One volunteer likens washing with Dry Soap to kissing a sea-cucumber. It may simplify the task of washing a baby, but it will sell best to outdoor, agricultural and military users, far from the decadent luxuries of piped water and hot towels.

David Jones

How old are distant stars?

SIR—The existence of young stars at vast distances from the plane of our Galaxy, as discussed in a recent News and Views article¹, seems to be reasonably well understood for some of the varieties of star under investigation. In the case of the young, high-velocity A stars observed in the galactic halo, however, evidence is now available that the identification of those stars as blue stragglers is incorrect.

My new observations of a large sample of A stars at the South Galactic Pole, from near the galactic plane to around 10 kiloparsecs away from the plane^{2,3}, give detailed information on their kinematics, abun-

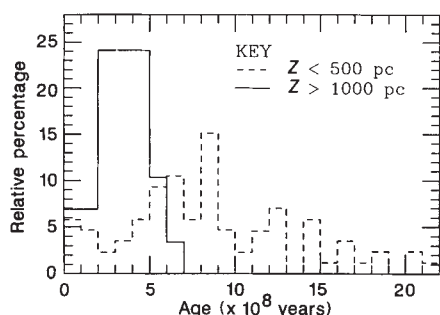


Fig. 1 The age distributions of two samples of young A stars at the South Galactic Pole, one group within 500 parsecs of the galactic plane, and the other at distances of 1–10 kiloparsecs away.

dances and, most importantly, their ages. The ages of main-sequence A stars may be derived using isochrones (same-age-evolution tracks) in colour–luminosity (H–R) diagrams for stars of similar abundances.

A collection of young (non-cluster) A stars will usually be randomly distributed on and above the main-sequence in an H–R diagram, because they are formed stochastically over time, and vary in

colour and luminosity as they evolve, at different rates, off the main-sequence towards the red-giant branch. Blue stragglers, descendants of some of these stars, seem to remain in ‘young’ locations for longer than normal lifetimes, so they also show the random distribution of their progenitors, plus their own evolutionary peculiarities (there is evidence that they often appear just above, rather than on, the main-sequence²).

The new data on A stars show that the age distributions for normal young A stars near the galactic plane, and for distant high-velocity A stars, are completely different. Those within 500 parsecs of the plane are randomly distributed, as expected, from recently formed stars up to those aged 2.2×10^8 years.

In striking contrast, the ages of young A stars at distances of more than 1 kiloparsec from the galactic plane are less than one-third of those observed for normal disk A stars — all are younger than $6.5 \pm 1.0 \times 10^8$ years old (Fig. 1). (This is not an effect of observational limits, because older A stars are intrinsically more luminous than main-sequence stars of the same colour, and if present, they would have been up to four times more likely to have been seen than the younger stars.)

Figure 2 shows the positions of some known blue stragglers in an H–R diagram, compared with those of the distant (1–10 kiloparsecs) A stars. In contrast to the random locations of the blue stragglers, the A stars have the classical appearance of a group of stars formed within a relatively short period, evolving away from the main-sequence. To propose that the A stars are actually blue stragglers is to suggest that a disparate collection of

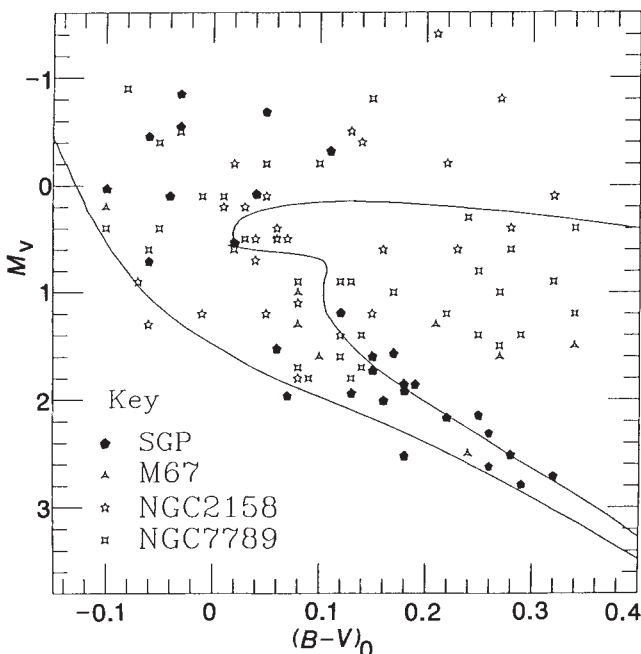


Fig. 2 The positions in an H–R diagram of young South Galactic Pole A stars more than 1 kiloparsec from the plane, compared with some known blue stragglers from the clusters M67, NGC2158 and NGC7789. Isochrones are shown for the main-sequence and for stars 5×10^8 years old.

anomalous stars, in existence for thousands of millions of years, sampled over a large volume of the Galaxy, would fortuitously mimic the distribution of a young coeval group. It is not necessary for us fully to comprehend the mechanism of blue-straggler (or distant A-star) formation to appreciate that, according to a standard classification technique like the H–R diagram, they are not the same kind of star.

Young A stars are seen far from the galactic plane in several directions³, which indicates that they are relatively widespread throughout our Galaxy. This distribution, plus their restricted age range, indicates that they originated from a single event that occurred on a galactic scale, such as the rapid accretion of a large quantity of gas, possibly from the merger of a satellite system with our Galaxy⁵. This would have resulted in the formation (from a mixture of accreted and disk gas) of high-velocity stars, at a specific epoch, with the unusual calcium abundances that are a feature of the distant young A stars^{2,3}.

It remains to be seen whether this particular hypothesis is correct. On present evidence, it is clear that blue stragglers, or any other type of star formed randomly over time from normal galactic processes, cannot be proposed to explain the existence of young high-velocity A stars far from the galactic plane.

CATHERINE M. LANCE

*Division of Radiophysics, CSIRO,
PO Box 76, Epping,
New South Wales 2121, Australia*

1. Trimble, V. *Nature* **336**, 111 (1988).
2. Lance, C.M. *Astrophys. J. Suppl. Ser.* **68**, 463–485 (1988).
3. Lance, C.M. *Astrophys. J.* **334**, 927–946 (1988).
4. Hagen, G.L. *Publ. David Dunlap Observ.* **4**, 3–62 (1970).
5. Rodgers, A.W., Harding, P. & Sadler, E. *Astrophys. J.* **244**, 912–918 (1981).

More insect eating

SIR—I would like to add some first-hand information to that provided by Sachi Sri Kantha on the subject of insectivory in Japan (*Nature* **336**, 316; 1988).

As a child in the Nanshin district of the southern Nagano prefecture in Japan, I consumed various species of insect. Many people living there still eat insects of ten subfamilies: Ephemeroptera (I cannot make further identification), Odonata (Libellulidae), Orthoptera (Tettigoniidae and Acridioidea), Plecoptera, Hemiptera (Cicadidae), Neuroptera, Tricoptera, Lepidoptera (pupae and adults of silk worm), Coleoptera (adults of true water beetles and water-scavenger beetles, plus larvae, pupae and adults of longicorn beetles) and Hymenoptera (all species of *Vespa* and *Polistes* inhabiting the district). In Nanshin supermarkets one can buy aquatic insects (canned), *Vespa* spp. (canned), rice hoppers and cooked pupae and adult silk worm. Most insects are cooked before eating but some people

eat raw larvae or pupae of wasps.

Kantha suggested insectivory could both provide a source of animal protein and remove agricultural pests. But another possibility is that insects are a source of salt and minerals. Nanshin is surrounded by mountains 2,000–3,000 metres high, and has heavy annual rainfall. These conditions must have made it difficult to bring in salt from coastal districts when traffic networks were still poorly developed.

A teacher at a high school in Nanshin tells me that until about 100 years ago, when the production of salt was reduced in coastal districts due to humid weather conditions, the people of Nanshin frequently either ate more insects to increase their salt intake or extracted salt by boiling *tatami*.

KATSUYA ICHINOSE

*The Institute of Low Temperature Science,
Hokkaido University,
Sapporo 060, Japan*

Superfamily expands

SIR — We report that two proteins of Epstein–Barr virus (EBV) are members of the immunoglobulin superfamily. The proteins were identified as immunoglobulin superfamily members by a program that can search for immunoglobulin-like domains in protein sequences. The similarity of the domains is mainly in the regions around two cysteines which form the conserved disulphide bond^{1,3}.

The algorithm of the program is the pattern-match method of Harr *et al.*⁴. First, two fragments that consist of the 16 amino-acid residues to the right of the first disulphide bond-forming cysteines and the 16 residues to the left of the second disulphide bond-forming cysteine of 49 representative immunoglobulin-like domains were tabled (data not shown). We wrote a statistical program to calculate the frequencies of 20 residues in each position of these fragments.

Based on those data, we wrote another program that could find immunoglobulin-like domains in Pascal and used it to search the PIR database (NBRF release 11.0) at a significance value of 10^{-16} . Of 3,606 sequences searched, we identified 279 known members of the immunoglobulin superfamily together with six previously unrecognized members. Two of these are EBV proteins.

One is the hypothetical BARF1 protein (NBRF code QQBE48), which is encoded by an open reading frame in *BamHI-A* fragment of the virus genome⁵. The other (QQBE4L) is a probable glycoprotein. These proteins have significance values of 9.8×10^{-16} and 8.9×10^{-15} , respectively. Local homology searches and secondary structure prediction with IDEAS programs¹⁶ confirm that residues 13–124 of QQBE48 and residues 20–135 of QQBE4L are similar to immunoglobulin

```

      40      50      60      70      80
QQBE4L VNLTCVSPPS-NEVSRIELGRVTPGQGLPLAVATSNNGTHNGGVNY
MHHUOU LTLTTFSGFSLTSSSRVSWIRRPQGALEULARIIBBQKFYSTSLRT
      90     100     110     120     130
QQBE4L SLTLEUVNDSTVSLSLIPNVT-LAHAGVYTCNVTLRNCSVASGVHCNY
MHHUOU RLKIS-KNDSHNDVLIIRINUPUTATAYV-BAKU-NSVMAGVYVYV

```

```

      20      30      40      50      60
QQBE48 VAAGQAVTAFLGERVTLTSYURKSLRPEIEVSUFLKGPGEF-QVLI
LZHUVL LTPFASVSSGSLGGFISCTGTSSDVSQVNVYVSGQGHATAPKLI
      70      80      90     100
QQBE48 GRMHQVIFIEWPFGRGFDIHRSAFTFLVVTAAINSHQSNYLIC
LZHUVL SEVRNPSGVS-D-R-FSGS-KSANTASLTISGLAEDAEADVVE

```

Two fragments of QQBE4L and QQBE48 aligned with two immunoglobulin variable regions (NBRF codes MHHUOU and LZHUVL). Identical residues (:) and conservative replacements (.) are highlighted. QQBE48 and QQBE4L are related to another 18 and 11 members of the superfamily, respectively, with s.d. values above 3. Top, homology score = -70, per cent match = 26.3, s.d. = 5.865. Bottom, homology score = -79, per cent match = 23.1, s.d. = 6.285.

variable regions and are predicted to have an immunoglobulin-fold structure (see figure). Furthermore, A. Williams (personal communication) has suggested that QQBE48 has a second domain (residues 126–221) distantly related to immunoglobulin constant domains.

The functions of the two proteins are unknown, and their immunoglobulin-like domain is unexplained. It is possible that EBV has at some stage integrated an immunoglobulin gene from its host B cell into its genome. Alternatively, one or both of the proteins may adhere to major histocompatibility complex molecules (or other molecules having an immunoglobulin-like domain) on the B-cell surface to increase the infectivity of the virus. This would resemble the way that most members of the superfamily perform their functions, such as cell recognition and adhesion, by means of associations of their immunoglobulin-like domains¹.

A case of a boy who suffered from chronic active EBV infection with Kawasaki-like disease was reported recently⁷. The patient's peripheral lymphocytes were EBV-receptor (CD21) negative, but could be infected by EBV. It will be interesting to see if QQBE48 or QQBE4L can bind to the CD4 molecule of the boy's T cells and cause the infection.

HUAICHUN WANG

JAJIN WU

PEIHSIEN TANG

*Institute of Basic Medical Sciences,
PO Box 130 (3), Beijing 100850,
People's Republic of China*

- Williams, A. F. & Barclay, A. N. *A. Rev. Immun.* **6**, 381–405 (1988).
- Barker, W. C., Ketcham, L. K. & Dayhoff, M. O. in *Atlas of Protein Sequence and Structure*, Vol. 5 (ed. Dayhoff, M.) Suppl. 3 (Natl. Biomed. Res. Found., Washington, DC, 1978).
- Matsunaga, T. & Mori, N. *Scand. J. Immun.* **25**, 485–495 (1987).
- Harr, R., Häggström, M. & Gustafsson, P. *Nucleic Acids Res.* **11**, 2942–2957 (1983).
- Baer, R. *et al. Nature* **310**, 207–211 (1984).
- Kanehisa, M. (ed.) *IDEAS-Integrated Databases and Extended Analysis System for Nucleic Acids and Proteins User Manual* (Natl. Cancer Institute, Frederick, Maryland, 1987).
- Kikuta, H. *et al. Nature* **333**, 455–457 (1988).

Lead contamination

SIR—Markovac and Goldstein (*Nature* **334**, 71–73; 1988) report that picomolar concentrations of lead stimulate protein kinase C. This conclusion may not be justified, as the authors do not seem to have taken into account the contamination of biochemical reagents with lead, and other metals. I have analysed a sample of MgCl₂ BDH Analar Grade (MgCl₂ was a component of their assay medium for PKC) by atomic absorption spectrophotometry and find 2×10^{-6} M of lead in a 1 M solution, equivalent to 1×10^{-8} M in a 5 mM solution. This is within the manufacturer's specification. Lead contamination can also be expected in other reagents used by Markovac and Goldstein, yet they report enzyme activation over a range from 10^{-16} to 10^{-12} M of lead. It is hard to believe that the lead concentration was actually in this range.

T.J.B. SIMONS

Department of Physiology,

King's College London,

Strand, London WC2R 2LS, UK

MARKOVAC AND GOLDSTEIN REPLY—Simon's concerns about reagent impurities are valid and, indeed, we did receive assurance from Sigma, from whom we purchased all our reagents except for 1,2-diolein (P-L Biochemicals), that they were lead-free at the sensitivity of the manufacturer's assay system.

The buffer used for protein kinase C preparation contained chelating agents to remove pre-existing cations, including calcium and lead. These chelators, present in trace amounts in our reaction mixture, should bind residual lead from reagent contamination. Because we did not perform atomic absorption analysis, we cannot be certain of the actual concentration of available lead, but the addition of 10^{-10} M lead to our basal assay system reproducibly stimulated protein kinase C activity, suggesting this important regulatory enzyme is highly sensitive to lead.

JASNA MARKOVAC

GARY W. GOLDSTEIN

Departments of Pediatrics and

Neurology,

University of Michigan Medical

School,

Ann Arbor, Michigan 48109-0570, USA

Selective membrane correction

Cd²⁺ should have been printed instead of Cu²⁺ on lines 3 and 13 of the third paragraph of the reply by Rubinstein *et al.* in *Nature* **337**, 217; 1989.

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of less than 500 words and five references. □

The ten commandments

David Weatherall

Genethics: The Clash Between the New Genetics and Human Values. By David Suzuki and Peter Knudtson. *Harvard University Press: 1989. Pp. 384. \$25.*

"GENETIC scientists patent terminally ill mice." "The 6 million Jews who died in concentration camps... were not innocent victims of the most cynical and systematic exercise in human degradation the world has yet seen; simply forerunners of the 'brave new world' order, apparently." "My company owns chromosome 7." The usually balanced sources of these recent quotes — a front-page headline in the *Independent* newspaper, correspondence in the *Radio Times* following a BBC television programme on the future of human genetics, and an article in *Science* — suggest that all is not well with human molecular genetics, or at least with the public's perception of what is happening in this field.

Why such bad press? How well are scientists presenting their case for the considerable medical potential of genetic engineering? In particular, are they succeeding in developing a properly informed public debate about its benefits and potential hazards? One of the problems in generating discussions of this type is that, to be of any value, both sides must understand the scientific, medical and ethical issues involved, at least in outline. With a few notable exceptions, the media, particularly in Britain, are making very little effort to explain the fundamentals of molecular biology or to present a balanced view of its medical possibilities. Rather the journalistic tendency is to concentrate on the more sensational and futuristic aspects of our increasing ability to manipulate the human genome. This may account for the recent plethora of books that attempt to explain the basics of molecular biology and to deal with some of the ethical problems of its application to human genetics. *Genethics* is the latest of this genre.

Like most of its fellows, *Genethics* falls into two parts. The early chapters provide a simple description of genetic principles and the way that it is now possible to dissect and analyse the human genome. Although some of their titles — "The Darwinian Dance", "The Dances of the Chromosomes", "Recombinant DNA: the New Choreography", for example — are irritatingly Wellsian (Sadlers rather than H.G., whose turn comes later), these chapters are written in a way that should enable newcomers to the field to grasp the basics of the new genetics.

It is the second part of the book that sets it apart from its predecessors. Here, each

chapter starts with what is called a "genethic principle". The first nine are fairly uncontroversial: the privacy of the individual genome; the restriction of genetic manipulation to somatic cells; the unacceptability of biological warfare; the inadvisability of crossing evolutionary boundaries in genetic manipulation; and so on. But the tenth — disclosed, like the identity of an Agatha Christie killer, only on the penultimate page — purports to provide the answer to all our problems. Its message is that, to search for ethical principles to guide us through the personal and collective decisions that stem from the applications of modern genetics, we will have to move from the "rigid boundaries of western science and philosophical thought", to the "rich, cross-cultural realms that embrace other ways of knowing". Apparently this means that the only way out of our dilemma is to seek help from the philosophies of oriental religion, the sacred texts of Islam or the "world-views" of American Indians.

Because I found the word "genethics" was becoming increasingly difficult to live with while this book was being reviewed, I turned for comfort to those usually reliable old friends, Partridge (*Usage and Abuse*) and Fowler (*Modern English Usage*). For once they let me down; both are fairly relaxed about neologisms. Fowler is happy if the new word is coined for brevity or poignancy; Partridge only requires some regard to etymological decency. "Genethics" certainly gets top marks for poignancy, though it scores less well as the illegitimate offspring of its Roman and Greek parents. But even if we accept the dubious assumption that this topic is so different from the rest of medical ethics that it warrants a new name, why the ten genethic principles? Are human geneticists to kneel and repeat them every night before going to bed? No doubt as they do they will reflect on the deeper meaning of the fact that there are ten of them. But just in case readers miss the point, Suzuki and Knudtson modestly remind them that their ten moral principles cannot claim to be either divinely inspired or cast in tablets of stone. But even for those (I suspect few) who can stomach this ritualistic approach there remains a problem. How valid, or helpful, is the authors' thesis that the ethical problems raised by genetic manipulation are beyond the reach of Western religion and philosophy, let alone common sense?

There has always been a fear of anything new in human biology, particularly if it affects medical practice. It is only necessary to think back a few years to the extraordinary public reaction to the first successful heart transplant. And yet such operations are now commonplace and publicly acceptable. The new approaches to prenatal diagnosis of genetic disease and somatic gene therapy, so well described in this book, raise no insurmountable ethical issues; they are simply extensions of what is being done already in day-to-day clinical practice. Now that somatic gene transfer seems to be on the horizon, sensible guidelines for all aspects of gene therapy have been drawn up, both in the United States and Europe. Even in that most sensitive of areas, embryo research, it is beginning to look as if we shall be able to reach a consensus about control and monitoring, though the regulations may well differ from country to country depending on the social, religious and cultural climates concerned. It is even possible that the legal profession will begin to sort out some of the current nonsense caused by slapping patents on every newly determined human DNA sequence. Although it is often a long, inefficient and painful process, once difficult ethical issues have been clearly defined and understood we usually manage to reach agreement on them.

So is there any point in getting all het up about genetic manipulation in the vague futuristic context of a Huxley or an Orwell, particularly as this is unlikely to be feasible in the foreseeable future and at a time when we have so many more tangible problems to tackle? Perhaps not, but many readers will feel sympathy for the authors' concerns, even if they have reservations about the way in which they are expressed.

Undoubtedly, much of the current unease about human genetic manipulation stems from the 'slippery-slope' argument. At the moment we are contemplating only the replacement of one defective gene among the 100,000 or so that constitute the human genome. A whole organ transplant is fine; why all the fuss about a few thousand nucleotide bases? If this turns out to be safe, and if it is acceptable to society, why not go the next step and correct a few genetic diseases by germ-line therapy? The family (and society) will be rid of the bad gene for good and all. And when we know more about the genes for traits that we might like to see expressed in our children, what harm will there be in indulging in a little genetic touching up? And so on.

Furthermore, it is not just scientific historians who still remember the way in which some of the outstanding human geneticists of the 1930s were seduced by the eugenics movement. Is the current

generation any more trustworthy? And many thoughtful people dislike the increasingly reductionist view of human nature that is resulting from molecular genetics, particularly when they are told that, by the end of the century, there will be a dictionary of all their genes. What will be disclosed when they are stripped down to their last nucleotide base? But what is not made clear in *Genethics*, and in many other popularized versions of the new genetics, is the vast complexity of the whole thing, let alone the fact that we still have only the flimsiest idea of how a gene actually works.

Fear and mistrust are generated by lack of understanding. The overall tone of *Genethics*, as highlighted by its title and its message that the future potential of human genetics is far too complex to be considered in the framework of Western ethics, philosophy and good sense, will serve only to increase public concern. The

subject needs a more balanced presentation of its potential (and, in particular, its inherent complexity), a little more humility on the part of its protagonists, and better-informed public debate on the ethical issues that are real rather than hypothetical. Many of the more ill-defined questions raised by this book, important though they may be, must be left for future generations to sort out; we can only hope that they will be better informed than ours. Meanwhile, although the myths of the Quechua, Maya or Haida may be a source of comfort to the authors of *Genethics*, for the worried young couple in the genetic counselling clinic or the lay members of bodies that control genetic manipulation or embryo research, something a bit nearer to home is required. □

Sir David Weatherall is Nuffield Professor of Clinical Medicine in the University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK.

Round referencing

Gwyneth Tseng, Goff Sargent and Jack Meadows

The Science Citation Index, Compact Disc Edition. Institute for Scientific Information. See text for details of prices.

SINCE the early 1960s, the Institute for Scientific Information has produced citation indexes which list the references at the end of research papers. The assumption is that if a particular paper is central to people's research, then papers citing that paper will also interest them. Each year, the *Science Citation Index* (SCI) lists over 600,000 references published in more than 3,300 journals. Besides their application as an information retrieval device, citation indexes are now used for other purposes and have attracted particular attention as a way of providing research profiles of individuals or groups through the number of citations to their work. In various more or less sophisticated applications, the SCI is being employed as an indicator of research virility.

The Institute for Scientific Information has now put the SCI onto compact disk. The disk reviewed here was for a quarter of a year only (disks are issued quarterly, with a complete accumulation available at the end of each year). The material is produced on two CD-ROMs plus a floppy

disk and a printed manual. The Title Word Access disk is primarily concerned with source items — that is, the papers scanned in preparing the index — while the Citation Access disk covers the items the selected papers have cited.

For those who have the MS-DOS extension, installation of the CD-ROM, which adheres to the High Sierra format, is simple. It requires a microcomputer (IBM or compatible) with 640K RAM, 5 megabytes of free hard-disk space, MS-DOS 3.1 or later, and a compatible CD-ROM player. We used an Amstrad PC1640 with a 20-megabyte hard disk linked to an Hitachi CDR1603s player. With the floppy disk in the selected drive, the set-up program gives only a few options, with defaults normally selected by pressing the return key. The search software is copied to a subdirectory on the hard disk and a batch file is written to the root directory. With a CD-ROM in the player, typing "SCI" runs the package.

Searches on either CD-ROM are saved and can be incorporated in searches on the other disk. For example, a keyword search can be made for a particular topic on the Title Word Access disk and the names of authors of relevant articles noted. The disks are then changed, and references to these source articles can be traced on the Citation Access disk. The menu-driven search interface is straightforward — an initial menu, backed up by subsidiary menus, associates function

keys with the facilities available. But although easy to use, this approach can be frustratingly slow and repetitive.

The CD-ROM version of SCI is actually easier to read than the printed version, while the good use of colour is an advantage over versions available online from various commercial vendors. (The online edition, SCISEARCH, is available through Data-Star in Switzerland, Dialog in the United States, DIMDI in West Germany and the University of Tsukuba in Japan. The first two are compared in the accompanying table.)

Unlike either the online or printed versions, CD-ROM also provides a "related records" facility which identifies papers citing some references in common with the paper studied and thereby allows rapid retrieval of a series of related papers. But compared with online, searches combining large sets of references are slow; for example, a test search on addresses looked for all papers published by divisions of Shell in Canada. The disk contained 80 items under Shell and 6,861 under Canada. Matching these records produced five publications from Shell Canada, but took altogether six minutes to complete on our equipment.

The SCI has been produced on CD-ROM for 1986 and 1987, to allow back-searching. For new subscribers, the price for 1988 is \$10,200 (which includes the software and manual), and the same price will apply for 1989. This does not represent a saving over the printed version, but if a subscriber already takes the printed version the CD-ROM is available additionally at half price. Quarterly disks can be exchanged at the end of the year for annual accumulations. Also, unlike many databases on CD-ROM, the disks are bought outright, not leased, so they remain the property of the purchaser.

The CD-ROM has some drawbacks, especially in terms of delay times in operation, but has attracted generally favourable comments from the users on whom we have tested it. It was particularly commended for its ease of use, clarity of display, ability to trace related records, availability of additional search criteria, and ability to print and save selected items to disk. The drawbacks, such as listing only first cited authors, are shared with the printed and online versions. We think it will be most sensibly purchased for heavy local use involving fairly lengthy searches for recently published items. Searches over longer periods of time, or requiring complex matching of retrieval criteria, are best done online, whilst smaller searches, especially over a very long time-base, can be carried out satisfactorily with the printed version. □

Gwyneth Tseng, Goff Sargent and Jack Meadows are in the Department of Library and Information Studies, Loughborough University, Leicestershire LE11 3TU, UK.

Comparison between two online versions of the Science Citation Index

	Timespan	Updating	Connect charge (per hour)	Offline print/ record	Online print/ record
Data-Star	1984–	Weekly	\$95	\$0.23	\$0.31
Dialog	1974–	Fortnightly	\$57, \$153*	\$0.46	\$0.46

* The two prices quoted for Dialog are for subscribers and non-subscribers to the printed version.

Things are looking up and up

Stephen P. Maran

Pathways to the Universe. By Francis Graham-Smith and Bernard Lovell. Cambridge University Press: 1988. Pp. 239. £15, \$24.95.

IN *Pathways to the Universe*, two leaders of the post-war revolution in radioastronomy have joined forces to produce a comprehensive survey of astronomy for the educated general reader. Sir Bernard and Sir Francis explain that "having enjoyed the privilege of pursuing research in a new

photographs of the Moon's surface were rushed to a 1964 meeting of lunar experts. There are many such elements in the book, in which the authors also employ the clever device of beginning each chapter by describing an experience or observation that readers can try for themselves. Each chapter ends with information drawn from the most recent investigations; that on comets, for example, contains an excellent summary of the visit of the International Cometary Explorer to the Comet Giacobini-Zinner and of the subsequent space missions to Halley's Comet.

On occasion, the narrative lapses into a dry recitation of facts, but the reader is encouraged to resume the journey upon encountering such rewarding chapters as the lucid discussion of cosmology or the skilful interweaving of facts, history and graphical information on meteors and micrometeorites. *Pathways* does leave us spinning when it comes to the rotation of

the Earth, however; to profit from that section, one really needs to be a trained physicist or even a Cambridge graduate. A few usages can be debated, too: for example, "Galactic centre", treated (as in "Roman orgy") as a capitalized adjective followed by a noun with its initial letter in lower case, seems like an international convention that should not be ratified.

I found only one certain oversight: Alvan Clark (p.152) did not photograph the white dwarf star Sirius B in 1862; rather, it was then that his son Alvan G. Clark spotted the tiny star visually while testing a new telescope on the brilliant Sirius A, and exclaimed excitedly, "Why father, the star has a companion". With this new book the reader, too, has found a trustworthy companion. □

Stephen P. Maran, 3414 Turner Lane, Chevy Chase, Maryland 20815, USA, is Editor-in-Chief of *The Reference Encyclopedia of Astronomy and Astrophysics*, to be published by Van Nostrand Reinhold in 1990.

Oculomotoring

I. M. L. Donaldson

Movements of the Eyes, 2nd edn. By R.H.S. Carpenter. Pion, 207 Brondesbury Park, London NW2 5JN: 1988. Pp. 593. £59.50, \$115.

OUR eyes are mobile and our perception of the visual world is built up as a series of snapshots, each taken with the eyes pointing in a slightly different direction. Eye movement is a pre-condition of seeing, and the relation between ocular displacement and visual perception presents problems which have occupied philosophers and natural scientists for centuries.

The movements themselves are simple compared to those of the limbs, in that they involve a system of constant inertia and are confined to rotations without translation. They are also exquisitely controlled. Thus, the oculomotor system is one that should be relatively easy to understand, and so should be a good testing ground for our understanding of the general problems of motor control.

The first edition of *Movements of the Eyes*, published in 1977, has become a standard work; what then of its successor? Although it is more than 40 per cent longer, with all of the sections rewritten to reflect the work of the past decade, the arrangement and overall scope remain much the same as before. Of the three sections, the first describes the types of eye-movement (there are six in human beings), and the second provides an excellent account of the static and dynamic properties of the globe and its muscles, and of the anatomy of the neural pathways which lead to and from the eye-muscles.

The third section deals with the whole system, and presents, explains and illustrates the mathematical models which may be used to describe and predict the system's behaviour; this section also confronts squarely the formidable problems of trying to identify the neural circuits which may people the black boxes of the mathematical models. There is an excellent new chapter on oculomotor adjustments; here Carpenter deals with the necessity for continual re-calibration of the oculomotor system to take account both of imperfections in its performance and of changing circumstances (for example, growth at one time of life and ageing at another).

One of the features of the first edition was an appendix on linear systems analysis for the non-mathematician. This has been expanded and is now a first-class introduction to the subject, requiring no technical mathematical knowledge — it is a remarkable achievement, and a necessary one, because the revised text relies upon a much more formalized use of systems analysis than before. The last section, in which models of the oculomotor system are described, includes perceptive discussions of the value of such schemes and of the usefulness of a strictly quantitative approach to the complexities of the nervous system. The author delivers a timely warning against being seduced by the elegance of such models into according them independent existence.

This new edition is a worthy successor to Carpenter's previous book. All those interested in the control of movement or in vision are once again in his debt. □

I.M.L. Donaldson is Professor of Neurophysiology in the Department of Pharmacology, University of Edinburgh, 1 George Square, Edinburgh EH8 9JZ, UK.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Sir Bernard Lovell in the operations room of the 250-foot radiotelescope named after him. The sight, during wartime, of short-lived echoes on a radar screen inspired Lovell to employ radar to search for echoes caused by showers of cosmic rays, using borrowed equipment and a field at Jodrell Bank lent by the Botanical Department of the University of Manchester. Since then, radioastronomy has grown into a big science, not least at Jodrell Bank.

field, we are now impelled to tell the story to the best of our abilities". Their abilities are more than equal to the task, and the authors are particularly successful at portraying the areas in which they themselves have had especially strong interests — meteors, space missions and the search for extraterrestrial intelligence (Lovell), and pulsars and radio galaxies (Graham-Smith).

The book is enlivened by their personal reminiscences of historic moments in modern astronomy, for example the sensation caused when the first close-up

Are you looking for a journal dealing with the pure physical aspects of biology?

You have now found it...

The Journal of Biological Physics

An International Journal for the formulation and Application of Physical and Mathematical Models in the biological sciences

Edited by Terence W. Barrett

The Journal of Biological Physics is an international medium for the publication of original papers, letters, critical commentaries, review articles on fundamental theoretical and experimental research developments in biological physics.

- multidisciplinary
- sound refereeing
- unique forum for communication on applications of new approaches to biology
- no pagecharges

Topics include: molecular energy transfer dynamics, enzyme physics, nonlinear mechanisms of molecular energy coupling and transport, fundamental mechanisms of bioenergetics, sensory receptor principles, quantum biophysics, information processing, dielectrical phenomena, cooperativity in molecular function, biophysics from the solid state perspective, molecular biophysics, mathematical biophysics, sensory communication principles, quantum quasiperiodicity and chaos, bionetics, origins of life, automata networks, neural networks and cellular automata.

Subscription Information ISSN 0092-0606

1989, Volume 17 (4 issues)

Institutional rate:

Dfl. 246.00/US\$120.00 (incl. p&h)

Please write for your free sample copy.



**KLUWER
ACADEMIC
PUBLISHERS**

P.O. Box 322, 3300 AH Dordrecht,
The Netherlands

P.O. Box 358, Accord Station,
Hingham, MA 02018-0358, U.S.A.

The resurrection of recapitulation

Rudolf A. Raff

A Theory of the Evolution of Development.
By Wallace Arthur. Wiley: 1988. Pp. 94.
£14.95, \$34.95.

LARGELY because of the spectacular success of developmental genetics, the main problems of embryonic development have been well publicized and are seen as central to understanding biology. Yet we are still struggling to incorporate development as a key part of evolutionary theory. This is a curious problem, because even Darwin recognized that development provides information on evolutionary history. Later it was realized that there is a crucial mechanistic connection, in that development translates genotype into phenotype. Why have we not yet managed to build a theory of evolution that satisfactorily incorporates development as well as genetics?

Wallace Arthur suggests that evolution and genetics were readily fused because both fields had powerful unifying laws, whereas development has not been absorbed because it lacks such a unifying theory. *A Theory of the Evolution of Development* is an attempt to provide this missing theory, and to show how it allows an integration of development with evolution — a tall order indeed.

Arthur presents some novel ideas and a coherent point of view, and his book has the great virtue of being concise yet clear. It is divided into five sections. The first presents Arthur's theory of the morphogenetic tree, which constitutes a theoretical model for development. Subsequent sections consider interactions between this model and selection, mechanisms for the principal evolutionary transitions, relationships to other theories, and an application of morphogenetic-tree theory to the evolution of higher taxa.

The morphogenetic tree is a diagrammatic construct that represents ontogeny as a tree, with causal connections between hierarchical levels represented as the branches. The tree has a basal point representing an initial 'morphogenetic heterogeneity'. As the tree forks, along the axis of developmental time, further levels of heterogeneity are added. The entire course of an ontogeny can be depicted in this way, and evolutionary changes of various sorts are readily visualized. This model is extremely simple and understandable, and it allows predictions to be made. But heuristic elegance comes at a high cost. To achieve simplicity, the nature of developmental processes is ignored, and some very static (and demonstrably incorrect) assumptions

about development have to be built in.

The most serious explicit assumption is that genes acting early in development have larger effects on adult phenotype than those acting later. Such a view is not new, and it seems to make sense. But it is a gross oversimplification, and it leads to a number of misleading predictions about development and how development must evolve. Thus most evolution by this model must be by addition of new steps near the end of development, a view very much in line with Haeckel's classic concept of recapitulation. Conversely, according to the model, evolution in early development must be much rarer and involve macromutations to other viable states, that is "morphological windows". Arthur presents some thought-provoking ideas on how such events might occur.

The problem with this outlook is that it ignores the ugly fact that early development frequently differs radically in related organisms — early development in direct-developing frogs or sea urchins is quite divergent from that of species with typical larval development. Furthermore, the famous diagram of all vertebrates diverging from a common stage, the pharyngula, with a tail and gills, is true only as far as it goes. Earlier development, including gastrulation, is very different in frogs, birds and mammals. It is the later and much more complex pharyngula stage that is constrained. It may be true that initial states for strictly dependent hierarchical processes will be less likely to evolve, but are we ready to describe early versus late development so confidently in such terms?

Arthur acknowledges that the developmental theory presented in his book lies in the long and honourable tradition of Waddington and other twentieth-century theorists. There is no question but that an integration of development and evolution is vital, but the time for theoretical constructs such as those presented here is either past or not yet with us. Concepts such as the morphogenetic tree, or older ones like the epigenetic landscape, are of heuristic value, but they can never substitute for real experimental data. A detailed analysis of development in specific organisms such as *Drosophila* has been made possible by the application of the tools of molecular genetics. These same tools can also be applied to the variations in developmental controls between organisms, including those not often seen in laboratories. We can now ask experimental questions about the evolution of homoeotic genes, about heterochrony, about developmental constraints. It is a time less for grand theorizing than for empirical studies on how development evolves. □

Rudolf A. Raff is a Professor in the Department of Biology, Indiana University, Bloomington, Indiana 47405, USA.

Surface properties probed by second-harmonic and sum-frequency generation

Y. R. Shen

Department of Physics, University of California, Materials and Chemical Sciences Division, Lawrence Berkeley Laboratory, Berkeley, California 94720, USA

Optical second-harmonic generation and the related technique of infrared-visible light sum-frequency generation are extremely versatile tools for studies of many kinds of surfaces and interfaces. With the help of ultra-short laser pulses, they can be used to monitor surface dynamics and reactions with sub-picosecond time resolution.

OPTICAL second-order harmonic generation is an exciting new technique for studying surfaces and interfaces. Surface science plays an important role in disciplines ranging from physics and chemistry to biology and electronics, where reliable characterization of surfaces and interfaces is essential. Accordingly, many tools for surface analysis have been developed over the years but most are limited in their application¹. For example, electron scattering and diffraction, photoemission, Auger spectroscopy and mass spectroscopy can be operated only in a vacuum. Optical techniques can generally be applied to any interface accessible to light but they may not have enough surface specificity and sensitivity and can therefore be used only if the signal contributed by the bulk can be suppressed².

Recently, several laser techniques have been developed for surface probing³. Among these, second-harmonic generation (SHG) has received much attention because of its simplicity, surface specificity and versatility.

The SHG technique is sufficiently sensitive to detect less than a monolayer of molecules adsorbed on a surface. As a surface probe, it has several attractive features. Being a laser-excited coherent optical process, surface SHG is highly directional. It is therefore suitable for non-detrimental, *in situ* remote-sensing studies and can be used for monitoring surfaces in a real environment. Sub-picosecond laser pulses can be used for probing, giving the technique the potential for sub-picosecond time resolution and making it useful for *in situ* real-time probing of fast surface dynamics and reactions. Because laser excitation is

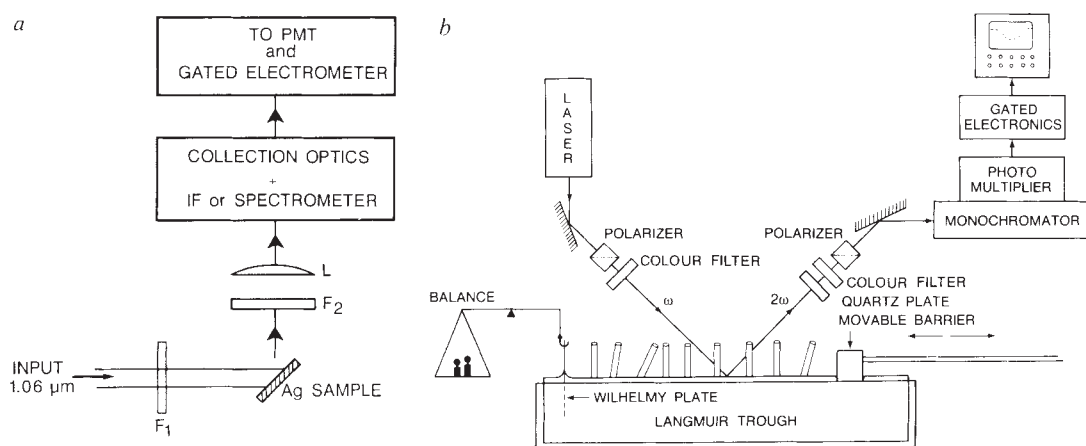
capable of high spatial and spectral resolutions, it also allows *in situ* mapping of the molecular arrangement and composition of a surface monolayer. These advantages, together with its general applicability to all types of interfaces, make SHG a unique and versatile tool for surface studies.

The SHG method is based on the principle that a second-order nonlinear optical process is forbidden in a medium with inversion symmetry, but such symmetry will necessarily be broken at an interface, where the nonlinear processes are therefore allowed. Consider an optical wave at frequency ω impinging on a medium. The wave can induce a dipole oscillation of each molecule in the medium. Because a molecule behaves like an anharmonic oscillator, overtone oscillations are excited at frequencies 2ω , 3ω and so on, as well as the fundamental at ω . Oscillating dipoles emit electromagnetic radiation, and hence one expects output radiation from the medium at frequencies ω , 2ω , 3ω and so on. In the SHG technique, one focuses attention on the frequency 2ω . For a driven anharmonic oscillator, the induced-dipole component, $p(2\omega)$, (which is the source of the second-harmonic radiation) is proportional to the quantity $E(\omega)E(\omega)$, the 'self product' of the incoming field of strength E , at frequency ω . The induced second-harmonic dipole per unit volume, $P^{(2)}(2\omega)$, can be written as

$$P^{(2)}(2\omega) = \chi^{(2)} E(\omega) E(\omega) \quad (1)$$

where $\chi^{(2)}$ is known as the nonlinear susceptibility and is a characteristic coefficient of the medium. If the medium has

Fig. 1 *a*, Schematic diagram of a typical experimental arrangement for surface second-harmonic generation. The incoming laser beam passes through a colour filter F_1 and impinges on the sample. The second-harmonic signal generated from the sample in the reflected direction is collected by the lens L , filtered through the colour filter F_2 and an interference filter (IF) or spectrometer, detected by the photomultiplier (PMT) and finally converted to an electrical signal by the gated electrometer. *b*, A setup used for second-harmonic generation (SHG) measurements of monolayers of molecules on a water surface. The Langmuir trough is filled with water on which molecules, shown as rods are adsorbed. The Wilhelmy plate hanging on the balance is used to measure the surface tension (π). Variation of the molecular-surface area (A) by the movable barrier allows π - A isotherms to be determined together with the SHG measurements.



inversion symmetry, the incoming fields $E(\omega)$ and $-E(\omega)$ must induce in the medium dipoles of $P(2\omega)$ and $-P(2\omega)$ respectively. This, however, is not consistent with equation (1) unless $\chi^{(2)} = 0$, indicating that SHG is forbidden. At an interface, the inversion symmetry must be broken, and radiation at frequency 2ω is therefore no longer forbidden. Consequently, SHG is highly surface specific.

Theory

The sub-monolayer sensitivity of SHG was noticed in 1973¹⁰ but received little attention. It was rediscovered in a search for a better understanding of the phenomenon of surface-enhanced Raman scattering^{11,12}. Several researchers suggested that this scattering arises because of enhancement of the local field at points on a rough metal surface^{13,14}. Raman scattering can be regarded as a two-photon nonlinear optical process, in which one photon is absorbed and another emitted. It was therefore suggested that other nonlinear optical processes should be similarly enhanced on rough metal surfaces¹⁵. Indeed, with a 1.06- μm laser pump, SHG is enhanced by a factor of $\sim 10^4$ on a roughened silver surface¹⁵. The signal is so strong that sub-monolayers of adsorbates on roughened silver can be monitored with a continuous-wave (CW) diode laser¹⁶. An approximate estimate of the signal strength indicates that even without surface enhancement, it should be possible to detect SHG from a molecular monolayer.

The theory of surface SHG has been well developed⁴⁻⁷, following the early work of Bloembergen and Pershan⁸, and we will discuss it only briefly. The physical origin of the surface non-linearity arises from two sources^{6,7}: structural discontinuity and field discontinuity at the interface. Which is the more important depends on the structure of the interfacial system. Structural discontinuity clearly dominates at a semiconductor surface with dangling bonds, for example, or where a monolayer of polar molecules is adsorbed with a preferred orientation at an interface. Field discontinuity, on the other hand, may dominate on

a liquid or glass surface, where the surface structure is not very different from that in the bulk. When higher-order effects are taken into account, the bulk should also contribute to the second-harmonic signal but it can be shown that the contribution is usually smaller than that from the surface. In any case, a surface modification should certainly be discernible by surface SHG even in the presence of a significant bulk contribution.

The strength of surface SHG can be estimated from equation A-2 (ref. 9; see panel below) typically, $|\chi_s^{(2)}|$ is of the order of 10^{-15} e.s.u. (here subscript s denotes a surface property). Assuming $L \approx 1$, $\epsilon \approx 4$ and $\sec^2 \theta \approx 2$, for the parameters in equation A-2, a laser beam with intensity $I(\omega) \approx 10 \text{ MW cm}^{-2}$, cross-sectional area $A \approx 0.2 \text{ cm}^2$ and pulse duration $T \approx 10 \text{ ns}$ at a wavelength of 1.06 μm should generate a second-harmonic output of $\approx 10^3$ photons per pulse. Signal strength is readily detectable with a photomultiplier, and points clearly to the sub-monolayer sensitivity of the technique. Equation A-2 indicates that the sensitivity should increase with laser intensity. Sensitivity is, however, eventually limited by the onset of surface damage at high beam intensity; to avoid surface damage or perturbation, the laser energy per unit area must be kept below a certain threshold value.

Applications

The experimental arrangement for surface SHG is fairly simple (Fig. 1). In general, pulsed lasers are used for excitation but CW lasers can also be used in some cases. The large difference in frequency between the fundamental (ω) and the second harmonic (2ω) makes filtering of the signal against unwanted background straightforward. Because the output is highly directional, spatial filtering can be used effectively to block most of the fluorescence or stray light from the sample area.

In the following, we discuss applications of surface SHG to various disciplines in surface science and give examples to illustrate the potential of the technique.

Metal surfaces. One of the most flourishing areas of research in

Surface SHG is an electromagnetic phenomenon governed by Maxwell's equations. The source of the output radiation at frequency 2ω is a surface polarization

$$\mathbf{P}_s^{(2)}(2\omega) = \chi_s^{(2)} : \mathbf{E}(\omega) \mathbf{E}(\omega) \quad (\text{A-1})$$

induced by the incoming laser field $\mathbf{E}(\omega)$ in the interface layer, where the surface susceptibility tensor $\chi_s^{(2)}$ is a characteristic constant of the interface layer. This corresponds to a sheet of coherently driven dipoles oscillating at 2ω . Like an array of antenna, these oscillating dipoles radiate coherently in a well defined direction. Solution of Maxwell's equations with $\mathbf{P}^{(2)}(2\omega)$ as the source term yields a second-harmonic output $S(2\omega)$ in the reflected direction (assuming an air/solid interface) given by

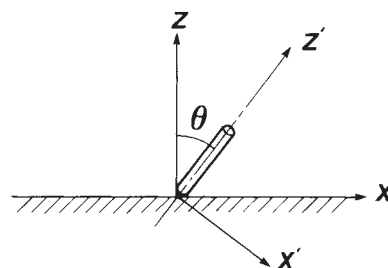
$$S(2\omega) = [32\pi^3 \omega \sec^2 \theta / \hbar c^3 \epsilon(\omega)^{1/2} (2\omega)] |L(2\omega)|^2 \times \chi_s^{(2)} : \mathbf{L}(\omega) \mathbf{L}(\omega) I^2(\omega) A T \text{ photons per pulse} \quad (\text{A-2})$$

where θ is the angle of incidence of the incoming laser beam, $\epsilon(\omega)$ is the dielectric constant at ω , $L(\omega)$ is a quantity that takes into account the Fresnel factor, and the beam polarization at ω , and $I(\omega)$, A and T are the intensity, the beam cross-section at the interface and the pulsewidth of the laser beam, respectively.

In equations (A-1) and (A-2), $\chi_s^{(2)}$ is a third-rank tensor which, in Cartesian coordinates, is defined by a set of 27 tensor elements $(\chi_s^{(2)})_{ijk}$, with i, j and k representing x, y or z . For example, according to equation (A-1), the incoming field components E_j and E_k can induce a surface-polarization component

$$(P_s^{(2)}(2\omega))_i = (\chi_s^{(2)})_{ijk} E_j(\omega) E_k(\omega) \quad (\text{A-3})$$

which is responsible for generation of the i -component of the second-harmonic field. Dictated by the symmetry of the interfacial system, many elements of $\chi_s^{(2)}$ can be vanishing or dependent on the others. With different combinations of input and output beam polarizations, the independent non-vanishing elements of $\chi_s^{(2)}$ can



The orientation of a rod-like molecule at an interface. The angle θ between the molecular axis z' and the surface normal z is measured by SHG using different combinations of input and output beam polarizations.

be deduced from the SHG measurements. In some cases, this also allows the average molecular orientation to be determined.

Consider the case where the molecule has a well defined long molecular axis and its nonlinear polarizability (or hyperpolarizability) tensor $\alpha^{(2)}$ is dominated by a single tensor element $\alpha_{z'z'z'}^{(2)}$ with z' along the molecular axis. One can easily relate $\alpha_{z'z'z'}^{(2)}$ to the various elements of $\chi_s^{(2)}$ in the laboratory coordinates (x, y, z) by a coordinate transformation. If θ is the angle between z' and the surface normal z (see figure) and the molecules have a random azimuthal distribution, then

$$\begin{aligned} (\chi_s^{(2)})_{zzz} &= N_s \alpha_{z'z'z'}^{(2)} \langle \cos^3 \theta \rangle \\ (\chi_s^{(2)})_{zxx} &= (\chi_s^{(2)})_{zyy} = (1/2) N_s \alpha_{z'z'z'}^{(2)} \langle \sin^2 \theta \cos \theta \rangle \end{aligned} \quad (\text{A-4})$$

where N_s is the surface density of the molecules and the angular brackets refer to an orientational average. From equation (A-4), it is seen that the measurements of $(\chi_s^{(2)})_{zzz}$ and $(\chi_s^{(2)})_{zxx}$ together with the assumption of a δ -function for the orientational distribution allow us to deduce both $\alpha_{z'z'z'}^{(2)}$ and θ .

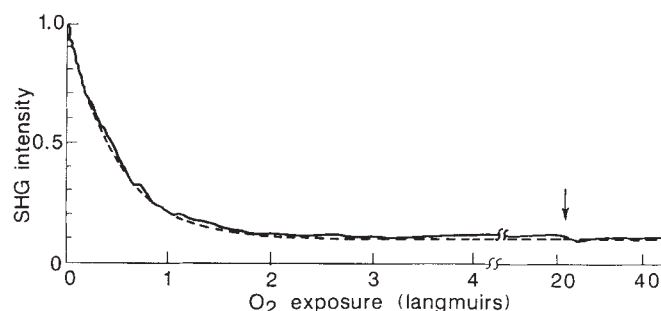


Fig. 2 SHG from a crystalline rhodium surface during O_2 exposure (1 langmuir unit = 10^{-6} torr second). The crystalline surface is specified by the Miller indices (1, 1, 1), that is, a plane intersecting the three orthogonal crystal axes at points equidistant from the origin. A full monolayer of oxygen on Rh(111) is characterized by a (2×2) LEED pattern, observed when O_2 exposure reaches ~ 20 langmuir units, indicated by an arrow. The theoretical curve (dashed line) used to fit the experimental result (solid line) was derived from the simpler Langmuir kinetic model.

modern surface science is the study of metal surfaces¹. Extensive studies of surface properties and molecule-substrate interactions on well-defined metal surfaces have been carried out in ultra-high-vacuum (UHV) conditions using conventional surface probes. The SHG technique can be adopted for tasks such as probing the molecular adsorption and desorption, processes involved in all surface problems of practical importance. Figure 2 shows an example of SHG used in UHV conditions to probe molecular adsorption on a metal surface, in this case oxygen on rhodium¹⁷. The signal from the surface decreases as the surface exposure to oxygen increases. This is qualitatively understood in terms of the fact that the nearly free electrons of a metal often dominate the surface nonlinearity and the adsorption of oxygen tends to localize the free electrons, hence reducing the nonlinearity. Measurement of the surface coverage by oxygen can be calibrated by low-energy electron diffraction (LEED). A full monolayer of oxygen, characterized by a LEED pattern that indicates (2×2) epitaxy, is observed for an oxygen exposure of ~ 20 langmuir units ($= 20 \times 10^{-6}$ torr s). As it is known that the O_2 molecule dissociates to atomic oxygen on a rhodium surface the result illustrated in Fig. 2 establishes three important facts: SHG has a sub-monolayer sensitivity; it can monitor the adsorption of atomic species; and it can follow the variation of surface coverage with time. The experimental data in Fig. 2 can be well fitted by a theoretical curve (dashed line) derived from the simple Langmuir kinetic model for adsorption. The 'sticking coefficient' for oxygen on the rhodium surface can be derived from this fit.

Several general properties of the technique are revealed by similar measurements carried out on other adsorbate/metal systems. SHG is sensitive to molecular adsorption at different sites¹⁷. It correlates well with results from thermal desorption spectroscopy and can be calibrated against the latter in determining surface coverage by adsorbates¹⁸. More accurate measurements of SHG can, in some cases, reveal the coverage dependence of the sticking coefficients¹⁹. SHG can probe surface structural changes and phase transitions of adsorbate overlayers²⁰. Finally, it can provide information on surface electron resonances²¹⁻²³ and surface structure of metals²⁴. In the future, the ultrafast response capability of the technique can be exploited to monitor surface reactions and surface dynamics on metals.

Semiconductor surfaces. The probing of semiconductor surfaces and interfaces is of great relevance because it can provide information about molecular adsorption, surface structure and surface processes. As with metals, molecular adsorption on silicon and germanium can easily be observed with SHG^{25,26}. The dangling bonds on clean silicon and germanium are respon-

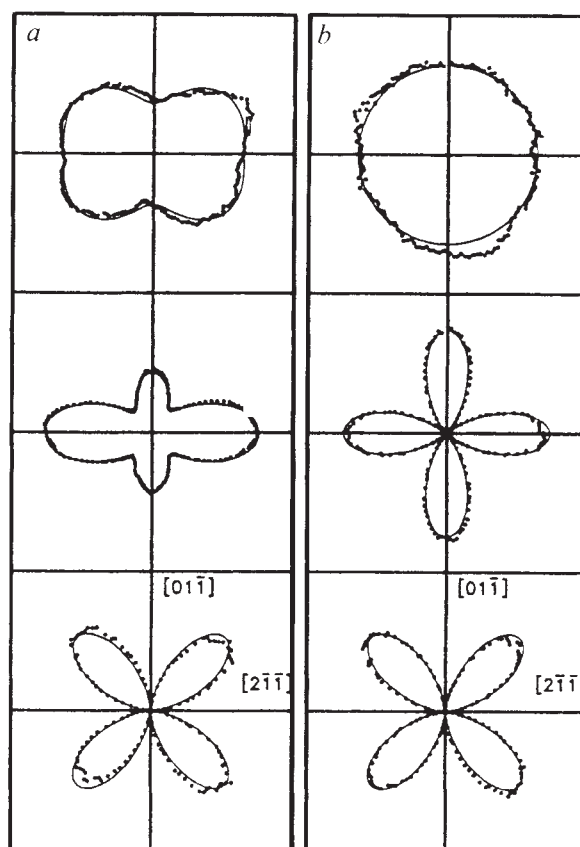


Fig. 3 Polar maps of SHG from *a*, a Si(111) surface characterized by a (2×1) LEED pattern, and *b*, a Si(111) surface characterized by a (7×7) LEED pattern; the polar angle denotes the direction of polarization of the normally incident pump beam. $[2\bar{1}\bar{1}]$ and $[01\bar{1}]$ specify two orthogonal directions in the crystal. The top panel displays the total second-harmonic signal; the middle and lower panels show, respectively, the signal polarized along $[2\bar{1}\bar{1}]$ and $[01\bar{1}]$. Dashed lines are experimental data and solid lines represent theoretical calculations.

sible for the rather large surface nonlinearities, because of the asymmetrical electron distribution at these bonds. Chemisorbed species that bind to the dangling bonds will reduce the asymmetry and hence the surface SHG. With $1.06\text{-}\mu\text{m}$ laser excitation, SHG from a clean silicon surface is completely dominated by the dangling bonds, so that the signal is nearly extinguished by adsorption of a full monolayer²⁴.

Surface structure governs the surface properties of a semiconductor. SHG can be used to probe the structural symmetry of a semiconductor surface layer and to monitor phase transitions *in situ*²⁶. This is reflected in the variation of the SHG signal as the sample or the beam polarization rotates about the surface normal²⁷. For example, a freshly cleaved Si(111) surface has structure characterized by a (2×1) LEED pattern, whereas the annealed surface has a structure characterized by a (7×7) LEED pattern. These two surface structures yield very different results of SHG as a function of rotation of the incoming beam polarization (Fig. 3)²⁶. On annealing, the freshly cleaved surface transforms from a (2×1) to a (7×7) structure; the process can be monitored *in situ* by the variation of SHG with time.

Other surface processes on semiconductors, such as surface melting²⁸⁻³¹, surface disordering³², amorphous-crystalline transformation³², formation of metal overlayers^{32,33} and change of surface electronic properties by adsorption (H. W. K. Tom and G. D. Aumiller, personal communication) can also be probed by SHG. Future applications should include monitoring of epitaxial growth, chemical vapour deposition, plasma deposition and etching, and growth of metal-semiconductor junctions.

The SHG technique as described so far does not work on semiconductors possessing no inversion symmetry, because the bulk contribution to the signal is overwhelming. It has, however, been demonstrated recently that in such cases, the bulk contribution can effectively be suppressed by polarization discrimination, thus regaining the surface sensitivity³⁴.

Liquid/solid interfaces. Although liquid/solid interfaces are generally considered as the most important area of surface science, the field has been little explored because of the lack of suitable tools. Surface SHG as a coherent optical technique seems ideal for providing information such as the free energy of adsorption and orientation of adsorbed molecules at interfaces. An example of the first is presented in Fig. 4, where the surface coverage of *p*-nitrobenzoic acid (PNBA) on fused quartz in an ethanol solution, as deduced from SHG, is plotted against the concentration of PNBA in the solution³⁵. Such a plot constitutes an adsorption isotherm, from which the free energy of adsorption for the system can be derived. PNBA molecules in the solution contribute little to the signal, confirming the surface-specific nature of the technique.

The orientation of molecular adsorbates at an interface reveals important information about their physical and chemical properties, and about the interaction between the adsorbates and the bonding medium. The molecular orientation at an interface, together with the nonlinear polarizability of the molecules, can sometimes be obtained from the surface SHG measurements (see panel). For the example shown in Fig. 4, the measurements yielded the elements of the surface susceptibility tensor $(\chi_s^{(2)})_{zzz} = 3.2 \times 10^{-16}$ e.s.u. and $(\chi_s^{(2)})_{xxx} = (\chi_s^{(2)})_{xyx} = 9.0 \times 10^{-17}$ e.s.u. for 0.53- μ m laser excitation. With a surface density of molecules $N_s = 2 \times 10^{14}$ cm⁻², these results yield $\alpha_{zzz}^{(2)} = 4 \times 10^{-30}$ e.s.u. and a tilt angle $\theta \approx 40^\circ$ (see panel and Fig. 1b). A similar experiment on PNBA at the air/fused quartz interface gave $\theta \approx 70^\circ$. The smaller value for θ for the liquid/solid interface presumably results from dielectric screening effect of the molecule-substrate interaction by the liquid.

Electrochemistry. Although possessing a long history, electrochemistry is poorly understood at the microscopic level. The enhancement of the SHG signal on roughened metal surfaces^{15,36} should make it easy to detect the appearance of a sub-monolayer of adsorbates on a metal electrode in an electrolyte and thus to monitor molecular adsorption and desorption in such systems.

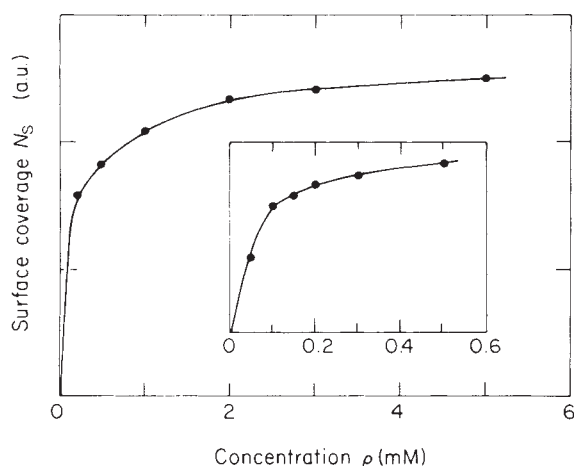


Fig. 4 Adsorption isotherm (surface coverage N_s plotted in arbitrary units as a function of volume concentration ρ) of *p*-nitrobenzoic acid (PNBA) adsorbed from ethanol solution by fused silica. The surface coverage of PNBA can be deduced from SHG under *s*-polarized excitation at 532 nm, as the SHG signal is proportional to the square of the surface coverage. The inset describes the data points at low concentrations. Coverage is in arbitrary units (a.u.).

Figure 5 shows a representative case, in which the surface SHG from a silver electrode in a 0.1 KCl solution is found to respond to the appearance and disappearance of AgCl on the electrode³⁷. In the oxidation-reduction cycle, the signal increases rapidly when the first monolayer of AgCl is deposited (as deduced from the amount of charge transfer through the electrode) and decreases rapidly when the last monolayer of AgCl is reduced. In between these two processes, hundreds of AgCl layers are deposited on and removed from the electrode, but the signal varies only slightly. This again demonstrates of the surface-specific nature of the technique.

Although SHG should have enough sensitivity to follow the adsorption of the first molecular monolayer on any electrode in any electrolytic process. Several different molecular species may coexist on the electrode and SHG is not selective enough to distinguish one from the others. Interpretation of the results then has to rely on identification of the species obtained by other methods.

In several recent experiments, the effective surface charge density resulting from ions adsorbed on an electrode has been deduced using the dependence of surface SHG on the bias voltage³⁹⁻⁴¹. The dynamics of the adsorption and desorption processes can be studied by time-resolved SHG measurements⁴². The technique has also been used to study electrodeposition of metal overlayers on an electrode in an electrolytic solution⁴²⁻⁴⁴, and to monitor the changes of surface symmetry⁴⁵ and surface properties of a crystalline electrode that result from a change in the bias voltage^{46,47} or from the deposition of metal overlayers^{44,48}.

Gas/liquid interfaces. Studies of molecular monolayers at gas/liquid interfaces can provide an insight into the understanding of the surface structure of liquids, surfactants, microemulsions and membranes, and the nature of wetting and two-dimensional phase transitions. Better control of Langmuir-type

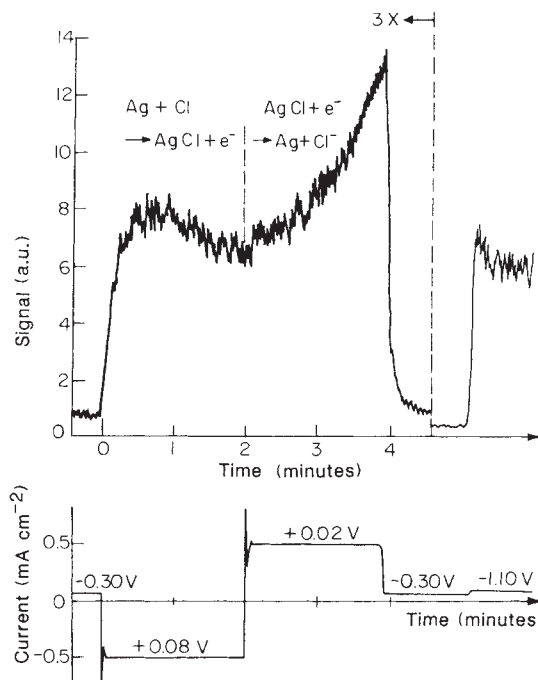


Fig. 5 SHG signal (top) and electrolytic current (bottom) from a silver electrode in a 0.5 M KCl solution as functions of time during an electrolytic cycle. The voltages listed on the lower curve indicate the potential of the silver electrode with respect to a saturated calomel electrode. The SHG signal is particularly sensitive to the initial deposition and final reduction of AgCl layers on the electrode. To the left of the vertical dashed line, the SHG signal is exaggerated by a factor of three.

molecular monolayers on liquid surfaces could yield high-quality Langmuir-Blodgett films, important in molecular-electronics applications. SHG is well established as an effective probe for gas/liquid interfaces and only a few demonstrations are described here.

SHG can be used to study phase transition in a monolayer of amphiphilic molecules on water. Measurement of the two-dimensional 'surface pressure' (π) as a function of surface area (A) of such a system often exhibits not only a two-dimensional gas-liquid phase transition but also, in some cases, a liquid expanded (LE)-liquid condensed (LC) phase transition. It was thought that the latter may be correlated with change in molecular orientation. As SHG can be used to determine molecular orientation, it is an ideal tool for this investigation. Such an experiment on pentadecanoic acid did indeed show a sudden orientational variation associated with the LE-LC transition⁴⁹.

The polymerization of a monolayer on water can also be monitored *in situ* with SHG. This is interesting because polymerization of a two-dimensional system is expected to be very different from the three-dimensional case. Moreover, the surface density of monomers can easily be varied, and the orientation of the monomers on water may be controlled by molecular synthesis. In such an experiment⁵⁰, a monolayer of octadecyl methacrylate monomers was spread on water and polymerized by irradiation of ultraviolet light. The surface SHG signal decreased accordingly and the time dependent results provided a check for existing theoretical models.

The rate constant of a surface reaction can be very different to that in the bulk and can be monitored by SHG. This has been clearly demonstrated in a recent experiment on a solution of *p*-nitrophenyl in water at various pH values⁵¹. Undoubtedly, SHG could also be used to probe liquid/liquid and solid/solid interfaces, but so far there are few examples. Orientations of amphiphilic molecules at different liquid/liquid interfaces, useful for the understanding of micellar or membrane structure, have been measured⁵².

Biological systems. SHG should soon find fruitful applications in biology. For example, molecules adsorbed on membranes tend to have a polar ordering, and should therefore yield a strong signal. SHG has been used to study adsorption of a monolayer of retinal chromophores on water, and on retinal chromophores first embedded in the purple membrane of *Halobacterium halobium* and then spread on water (J. Huang, A. Lewis, Th. Rasing, T. Stehlin and Y. R. Shen, unpublished results). The orientation, nonlinear polarizability and optical spectrum of the chromophores were obtained, and the change in the dipole moment of the molecules upon optical excitation could be deduced. Microscope images of biological tissues can also be obtained with SHG, providing structural information on the scale and form of molecular ordering⁵³. In this case, however, the signal was generated mostly from the bulk of the tissue.

Surface monolayer microscopy. SHG can be used to probe the morphology of a surface monolayer. This was demonstrated by the observation of the SHG image of a $\sim 8\text{-}\mu\text{m}$ laser-ablated hole in a Rhodamine 6G dye monolayer on fused quartz¹⁶. By using a focused laser beam, the spatial resolution is limited in principle only by the optical wavelength.

Sum-frequency generation.

With the help of a tunable laser, SHG can be used for surface spectroscopy⁵⁴. The underlying principle is straightforward: when either of the frequencies ω or 2ω is close to that of some molecular excitation, the nonlinear susceptibility $\chi^{(2)}$ should exhibit a resonant enhancement. Spectroscopic study is, however, limited to electronic transitions in the visible, because a high-gain photodetector, such as a photomultiplier, is needed to achieve monolayer sensitivity. To study adsorbed molecules selectively, infrared vibrational spectra are more useful because

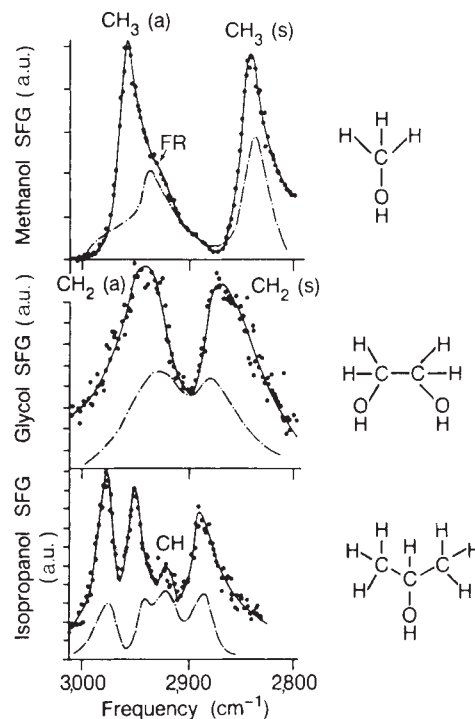


Fig. 6 Sum-frequency generation as a function of infrared input frequency for three adsorbed species, CH_3OH , $\text{C}_2\text{H}_4(\text{OH})_2$ and $\text{C}_3\text{H}_7\text{OH}$, on fused silica. The spectra were obtained with a visible input at $0.532\text{ }\mu\text{m}$ and a tunable input around $3\text{ }\mu\text{m}$. The peaks correspond to the various hydrocarbon (C-H) stretch modes, symmetric(s) and asymmetric(a); FR denotes a structure resulting from Fermi resonance between $\text{CH}_3(\text{s})$ and the overtone of the C-H bending modes. For comparison the Raman spectra of the species in the liquid phase (dashed-dot curves) are also shown.

they constitute 'fingerprints' for identifying molecules but appropriate photomultiplier is available in the infrared. An alternative method, known as sum-frequency generation (SFG), has recently been developed to overcome this problem⁵⁵⁻⁵⁷.

Any effective nonlinear optical method for surface vibrational spectroscopy should satisfy three criteria: it should be a second-order process (that is, responsive to anisotropy) so that it is surface-specific; input must have a tunable infrared component to excite vibrational transitions; and the output should be in the near-infrared or visible, where it can be detected by a photomultiplier. Infrared-visible SFG is an excellent candidate and the technique is a simple extension of SHG. The incoming field E is of two frequencies, ω_1 and ω_2 ; the output is then no longer at 2ω but at $\omega_1 + \omega_2$, or originating from the surface nonlinear polarization $P_s^{(2)}(\omega_1 + \omega_2)$ induced by the incoming fields. If ω_1 is in the infrared and ω_2 in the visible, all three of the above criteria are satisfied by SFG. Thus, SFG shares the surface specificity of SHG and in addition has capable of selective detection of molecules through their characteristic vibrational transitions.

Surface vibrational spectroscopy by SFG has been applied to the study of molecular adsorbates at various interfaces. Figure 6 shows the SFG spectra of C-H bond stretching vibrations of three different molecular species adsorbed on fused quartz⁵⁸. The different stretching modes are clearly resolved and the species can be distinguished by their spectra. The example depicted in Fig. 7a confirms the surface sensitivity of the method. Here the SFG spectra of C-H stretches are plotted for three liquid/solid interfacial systems: hexadecane/silica, hexadecane/OTS/silica and CCl_4 /OTS/silica, where OTS denotes a full monolayer of octadecyltrichlorosilane adsorbed on silica⁵⁹. Because surface SFG, like SHG, receives little contribution from

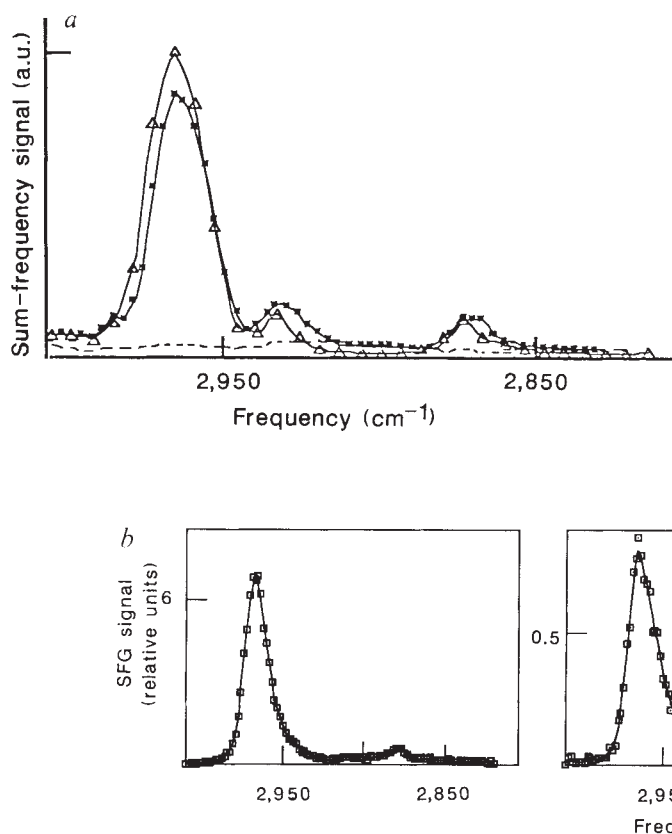


Fig. 7 *a*, SFG spectra of different interfaces obtained with a $\hat{p}$ -polarized visible input at $0.532\ \mu\text{m}$ and a $\hat{p}$ -polarized tunable visible input around $3\ \mu\text{m}$. Dashes: hexadecane/silica; solid squares: hexadecane/OTS/silica; triangles: CCl_4 /OTS/silica; OTS denotes a monolayer of octadecyltrichlorosilane adsorbed on silica. The solid lines are guides for the eye. OTS, when present, dominates the spectra. *b*, Sum-frequency spectra at the air/OTS/silica interface for various polarization combinations of the visible ($0.532\ \mu\text{m}$) and infrared input beams. Left to right, visible: $\hat{p}$ -polarized, infrared: $\hat{p}$ -polarized; visible: $\hat{p}$, infrared: $\hat{s}$; visible: $\hat{s}$, infrared: $\hat{p}$. The peaks in the spectra can be assigned to C-H stretch modes of the terminal methyl group of OTS. The peak at $2,878\ \text{cm}^{-1}$ originates from a CH_3 s-stretch, that at $2,964\ \text{cm}^{-1}$ from a CH_3 a-stretch, and that at $2,942\ \text{cm}^{-1}$ from the Fermi resonance between CH_3 (s) and the overtone of the C-H bending mode. Peaks for the CH_2 stretches on the alkane chain of OTS are hardly visible for reasons of symmetry.

the isotropic bulk, the spectrum for the first system shows no visible peaks, even though liquid hexadecane should have a strong infrared absorption band in this spectral region arising from the large number of C-H stretches in these molecules. Comparison of the two spectra for CCl_4 /OTS/silica and hexadecane/OTS/silica also demonstrates the surface specificity: the two spectra are nearly identical, both exhibiting the stretch modes only of the terminal methyl group of OTS.

The polarization dependence of SFG spectra also yields useful information. Figure 7*b* shows the spectra of the air/OTS/silica interface for three different polarization combinations⁵⁹. The peak heights are clearly different in each case. All the peaks in the spectra are again associated with the C-H stretches of the terminal methyl group of OTS, and the polarization dependence results from the orientation of this group. A quantitative analysis can be performed to calculate this orientation. In Fig. 7*b*, the data yield an angle of $\sim 45^\circ$ between the CH_3 symmetric axis and the surface normal, indicating that, as expected, the alkyl chain of OTS is nearly parallel to the surface normal.

SFG surface vibrational spectroscopy has also been applied to metal and semiconductor surfaces^{60,61}. The use of sub-picosecond pump pulses further permits the study of surface dynamics of selected molecular species. The fast response of SFG may even provide a means for following selective surface reactions and for probing intermediate species in the reactions. At present, the main difficulty of this technique is the lack of tunable sources in the mid- and far-infrared range but this problem should be overcome when infrared free-electron lasers become generally available.

In summary, SHG is an extremely versatile and valuable tool for surface studies that can be applied to all interfaces accessible to light and has many unique advantages over conventional surface probes. In particular, its sub-picosecond time resolution offers possibilities for research on ultrafast surface dynamics and reactions. The lack of molecular selectivity is overcome by using SFG for surface vibrational spectroscopy. These nonlinear

optical techniques should soon be widely adopted for surface and interface studies in many disciplines.

This work was supported by the Direction, Office of Energy Research, Office of Basic Energy Sciences, Materials Sciences Division of the US Department of Energy.

1. Somorjai, G. *Chemistry in Two Dimensions: Surfaces* (Cornell University Press, New York, 1981).
2. Aspnes, D. E. & Studna, A. A. *Phys. Rev. Lett.* **54**, 1956-1959 (1985).
3. Hall, R. B. & Ellis, A. B. (eds) *New laser and Optical Investigation of Chemistry and Structure at Interfaces* (Deerfield Beach, Florida, 1986).
4. Bloembergen, N., Chang, R. K., Jha, S. S. & Lee, C. H. *Phys. Rev.* **174**, 813-822 (1968).
5. Wang, C. C. *Phys. Rev.* **178**, 1457-1461 (1969).
6. Guyot-Sionnest, P., Chen, W. & Shen, Y. R. *Phys. Rev. B* **33**, 8254-8263 (1986).
7. Guyot-Sionnest, P., Shen, Y. R. *Phys. Rev. B* **35**, 4420-4426 (1987).
8. Bloembergen, N. & Pershan, P. S. *Phys. Rev.* **128**, 606-622 (1962).
9. Shen, Y. R. *The Principles of Nonlinear Optics* Ch. 25 (Wiley, New York, 1984).
10. Chen, J. M., Bower, J. R., Wang, C. S., Lee & C. H., *Opt. Commun.* **9**, 132-134 (1973).
11. Fleischmann, M., Hendra, P. J. & McQuillan, A. J. *Chem. Phys. Lett.* **26**, 163-166 (1974).
12. Jeanmaire, D. & van Dyne, R. P. J. *electroanal. Chem.* **84**, 1-20 (1977).
13. Moskovits, M. *J. chem. Phys.* **69**, 4159-4161 (1978).
14. Gersten, J. I. & Nitzan, A. *J. chem. Phys.* **73**, 3023-3037 (1980).
15. Chen, C. K., de Castro, A. R. B. & Shen, Y. R. *Phys. Rev. Lett.* **46**, 145-148 (1983).
16. Boyd, G. T., Shen, Y. R. & Hansch, T. W. *Opt. Lett.* **11**, 97-99 (1986).
17. Tom, H. W. K. *et al. Phys. Rev. Lett.* **52**, 348-351 (1984).
18. Zhu, X. D., Shen, Y. R. & Carr, R. *Surf. Sci.* **163**, 114-120 (1985).
19. Grubb, S. G., DeSantolo, A. M. & Hall, R. B. *J. phys. Chem.* **92**, 1419-1425 (1988).
20. Heskett, D., Song, K. J., Burns, A., Plummer, E. W. & Dai, H. L. *J. chem. Phys.* **85**, 7490-7491 (1986).
21. Tom, H. W. K. *et al. Surf. Sci.* **172**, 466-476 (1986).
22. Hener, W., Schröder & Zacharias, H. *Chem. Phys. Lett.* **135**, 299-302 (1987).
23. Giesen, K. *et al. Physica Scripta* **35**, 578-581 (1987).
24. Tom, H. W. K. & Aumiller, G. D. *Phys. Rev. B* **33**, 8818-8821 (1986).
25. Tom, H. W. K., Zhu, X. D., Shen, Y. R. & Somorjai, G. A. *Surf. Sci.* **167**, 167-176 (1986).
26. Heinz, T. F., Loy, M. M. T. & Thompson, W. A. *Phys. Rev. Lett.* **54**, 63-66 (1985).
27. Tom, H. W. K., Heinz, T. F. & Shen, Y. R. *Phys. Rev. Lett.* **51**, 1983-1986 (1983).
28. Shank, C. V., Yen, R. & Hirlimann, C. *Phys. Rev. Lett.* **51**, 900-902 (1983).
29. Akhmanov, S. A. *et al. Opt. Commun.* **47**, 202-204 (1983).
30. Malvezzi, A. M., Liu, J. M. & Bloembergen, N. *Appl. Phys. Lett.* **45**, 1019-1021 (1984).
31. Tom, H. W. K., Aumiller, G. D. & Brito-Cruz, C. H. *Phys. Rev. Lett.* **60**, 1438-1444 (1988).
32. Heinz, T. F., Loy, M. M. T. & Thompson, W. A., *J. Vac. Sci. Technol.* **B3**, 1467-1470 (1985).
33. Iyer, S. S., Heinz, T. F. & Loy, M. M. T., *J. Vac. Sci. Technol.* **B5**, 709-710 (1987).
34. Stehlin, T., Feller, M., Guyot-Sionnest, P. & Shen, Y. R. *Opt. Lett.* **13**, 389-391 (1988).
35. Heinz, T. F., Tom, H. W. K. & Shen, Y. R. *Phys. Rev. A* **28**, 1883-1885 (1983).
36. Boyd, G. T., Rasing, Th., Leite, J. R. R. & Shen, Y. R. *Phys. Rev. B* **30**, 519-526 (1984).
37. Chen, C. K., Heinz, T. F., Ricard, D. & Shen, Y. R. *Phys. Rev. Lett.* **46**, 1010-1012 (1981).
38. Murphy, D. V., von Raben, K. U., Chen, T. T., Owen, J. F. & Chang, R. K. *Surf. Sci.* **124**, 529-546 (1983).
39. Corn, R. M., Romagnoli, M., Levenson, M. D. & Philpott, M. R., *Chem. Phys. Lett.* **106**, 30-35 (1984).

40. Richmond, G. L. *Chem. Phys. Lett.* **110**, 571-575 (1984).
41. Richmond, G. L., Rojhanalab, H. M., Robinson, J. M. & Shannon, V. L. *J. Opt. Soc. Am. B* **4**, 228-236 (1987).
42. Corn, R. M., Romagnoli, M., Levenson, M. D. & Philpott, M. R. *J. chem. Phys.* **81**, 4127-4132 (1984).
43. Furtak, T. E., Meragliotta, J. & Korenowski, G. M. *Phys. Rev. B* **35**, 2596-2572 (1987).
44. Richmond, G. L., Koos, D. A., Robinson, J. M. & Shannon, V. L., *Bull. Am. phys. Soc.* **33**, 1648 (1988).
45. Shannon, V. L., Koos, D. A. & Richmond, G. L. *J. chem. Phys.* **87**, 1440-1441 (1987); *Appl. Opt.* **26**, 3579-3583 (1987).
46. Shannon, V. L., Koos, D. A. & Richmond, G. L. *J. phys. Chem.* **91**, 5548-5555 (1987).
47. Shannon, V. L., Koos, D. A., Robinson, J. M. & Richmond, G. L. *Chem. Phys. Lett.* **142**, 323-328 (1987).
48. Miragliotta, J. & Furtak, T. E. *Phys. Rev. B* **37**, 1028-1030 (1988).
49. Rasing, Th., Kim, M. W., Shen, Y. R. & Grubb, S. *Phys. Rev. Lett.* **55**, 2903-2906 (1985).
50. Berkovic, G., Rasing, Th. & Shen, Y. R. *J. chem. Phys.* **85**, 7374-7376 (1986).
51. Bhattacharyya, K., Sitzmann, E. V. & Eienthal, K. B. *J. chem. Phys.* **87**, 1442-1443 (1987).
52. Grubb, S. G., Kim, M. W., Rasing, Th. & Shen, Y. R. *Langmuir* **4**, 452-454 (1988).
53. Freund, I. & Deutsch, M. *Opt. Lett.* **11**, 94-96 (1986).
54. Heinz, T. F., Chen, C. K., Ricard, D. & Shen, Y. R. *Phys. Rev. Lett.* **48**, 478-481 (1983).
55. Zhu, X. D., Suhr, H. & Shen, Y. R. *Phys. Rev. B* **35**, 3047-3050 (1987).
56. Hunt, J. H., Guyot-Sionnest, P. & Shen, Y. R. *Chem. Phys. Lett.* **133**, 189-192 (1987).
57. Guyot-Sionnest, P., Hunt, J. H. & Shen, Y. R. *Phys. Rev. Lett.* **59**, 1597-1600 (1987).
58. Hunt, J. H., Guyot-Sionnest, P. & Shen, Y. R. in *Laser Spectroscopy VIII* (eds Persson, W. & Svanberg, S.) 253-266 (Springer, Berlin, 1987).
59. Guyot-Sionnest, P., Superfine, R. & Hunt, J. H. *Chem. Phys. Lett.* **144**, 1-5 (1988).
60. Harris, A. L., Chidsey, C. E. D., Levinos, N. J. & Loiacono, D. N. *Chem. Phys. Lett.* **141**, 350-356 (1987).
61. Superfine, R., Guyot-Sionnest, P., Hunt, J. H., Kao, C. T. & Shen, Y. R. *Surf. Sci.* **200**, L445-L450 (1988).

ARTICLES

Designing CD4 immunoadhesins for AIDS therapy

Daniel J. Capon, Steven M. Chamow*, Joyce Mordenti†, Scot A. Marsters, Timothy Gregory*, Hiroaki Mitsuya†, Randal A. Byrn§, Catherine Lucas||, Florian M. Wurm†, Jerome E. Groopman§, Samuel Broder† & Douglas H. Smith

Departments of Molecular Biology, * Recovery Process Research and Development, † Pharmacological Sciences, || Medicinal and Analytical Chemistry, ¶ Cell Culture Research and Development, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California, 94080, USA

‡ The Clinical Oncology Program, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, 20892, USA

§ Division of Hematology-Oncology, Harvard Medical School, New England Deaconess Hospital, Boston, Massachusetts, 02215, USA

A newly-constructed antibody-like molecule containing the gp120-binding domain of the receptor for human immunodeficiency virus blocks HIV-1 infection of T cells and monocytes. Its long plasma half-life, other antibody-like properties, and potential to block all HIV isolates, make it a good candidate for therapeutic use.

DESPITE the exquisite ability of the immune system to distinguish between self and non-self, and to put forth an impressive diversity in its antigen-recognizing repertoire, it can still be outflanked by a rapidly changing pathogen. Human immunodeficiency virus type 1 (HIV-1) is an example of such a pathogen, and, as a result, its consequences are devastating. Every individual infected with the virus is expected to develop a serious or life-threatening illness¹; no protective state has been shown to be generated in natural infections. It has not yet been possible to generate a protective response by immunizing chimpanzees with gp120, the HIV-1 envelope glycoprotein^{2,3}, or to confer passive immunity to chimpanzees using human IgG⁴. Even neutralizing antibodies made in experimental animals can block the infectivity of only a few HIV-1 isolates^{3,5}. Thus, the prospects for eliciting protective immunity against HIV-1, or for using antibodies as therapeutic agents to control HIV-1 disease are bleak. Anti-retroviral chemotherapy using dideoxynucleosides such as AZT does help some patients, but the toxicity is such that new strategies are needed⁶.

We have therefore attempted to block HIV-1 infectivity with soluble derivatives of CD4, the receptor for HIV-1, with the rationale that the CD4-binding domain of gp120 is the only part of gp120 that the virus cannot afford to change⁷. CD4 is a cell-surface glycoprotein found mostly on a subset of mature peripheral T cells that recognize antigens presented by class II MHC molecules^{8,9}. Antibodies to CD4 block HIV-1 infection of T cells^{10,11} and human cells not susceptible to HIV-1 infection become so after transfection with a CD4 cDNA¹². Gp120 binds CD4 with high affinity ($K_D \sim 10^{-9}$ M), suggesting that it is this interaction which is crucial to the entry of virus into cells^{7,13}. Indeed, we⁷ and others¹⁴⁻¹⁸ have shown that soluble rCD4, lacking the transmembrane and cytoplasmic sequences of CD4, can block HIV-1 infectivity, syncytium formation, and cell killing by gp120 (ref. 19). rCD4 blocks the infectivity of diverse HIV-1 isolates (R.B., J.G., H.M. and S.B., unpublished results),

and in theory should block all. At best, however, soluble rCD4 offers only a passive defence against the virus.

Active immunity requires a molecule such as an antibody, which can specifically recognize a foreign antigen or pathogen and mobilize a defence mechanism. Antibodies comprise two functionally independent parts, a rather variable domain (Fab), which binds antigen, and an essentially constant domain (Fc), providing the link to effector functions such as complement or phagocytic cells. It is almost certainly the lack of an antigen-binding domain which can neutralize all varieties of virus that hampers the development of humoral immunity to HIV-1. We reasoned that the characteristics of CD4 would make it ideal as the binding site of an antibody against HIV-1. Such an antibody would bind and block all HIV-1 isolates, and no mutation the virus could make, without losing its capacity to infect CD4⁺ cells specifically, would evade it. We therefore set out to construct such an antibody by fusing CD4 sequences to antibody domains.

We had two major aims for our hybrid molecules; first, as pharmacokinetic studies in several species predict that the half-life of soluble CD4 will be short in humans (30-120 min; J.M., unpublished results) we wished to construct a molecule with a longer half-life; second, we wanted to incorporate functions such as Fc receptor binding, protein A binding, complement fixation and placental transfer, all of which reside in the Fc portion of IgG. The Fc portion of immunoglobulin has a long plasma half-life, like the whole molecule, whereas that of Fab is short, and we therefore expected to be able to fuse our short-lived CD4 molecule to Fc and generate a longer-lived CD4 analogue. Because CD4 is itself part of the immunoglobulin gene superfamily, we expected that it would probably fold in a way that is compatible with the folding of Fc. We have therefore produced a number of CD4-immunoglobulin hybrid molecules, using both the light and the heavy chains of immunoglobulin, and investigated their properties. We have named one

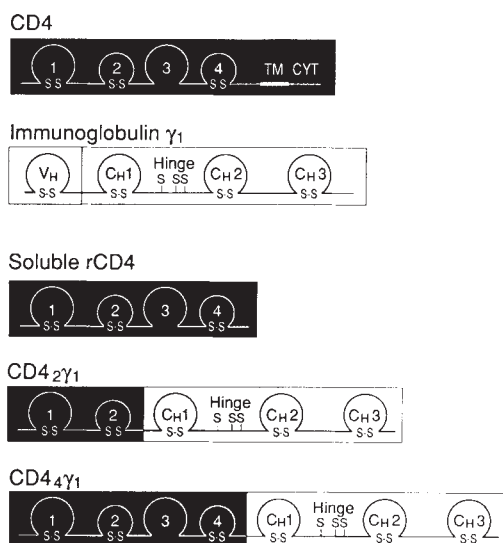


Fig. 1 Structure of cell surface CD4, human IgG1 (γ_1), soluble rCD4, and CD4 immunoadhesins ($2\gamma_1$ and $4\gamma_1$). The immunoglobulin-like domains of CD4 are numbered 1 to 4; TM and CYT refer to the transmembrane and cytoplasmic domains. Soluble rCD4 is truncated after proline 368 of the mature CD4 polypeptide. This results in a secreted, soluble polypeptide with an affinity for gp120 similar to that of cell surface CD4 (ref. 7). The vertical division within IgG1 indicates the junction of the variable (VH) and constant (CH1, hinge, CH2, and CH3) regions. Disulphide bonds formed within IgG1 domains and the immunoglobulin-like domains of CD4 are indicated by (S-S). The positions of cysteine residues that form intermolecular disulphide bridges connecting the IgG1 heavy-chain hinge to light and heavy chains are indicated by (S). CD4-derived and IgG1-derived domains of $2\gamma_1$ and $4\gamma_1$ are indicated by shaded and unshaded regions, respectively. The $2\gamma_1$ and $4\gamma_1$ immunoadhesins consist of residues 1 to 180 and residues 1 to 366 of the mature CD4 polypeptide, respectively, fused to the first residue (serine 114) of the human IgG1 heavy-chain constant region.

Methods. For the expression of CD4 immunoadhesins, the sequences of CD4 and human IgG1 were fused by oligonucleotide-directed deletional mutagenesis after their insertion into a mammalian expression vector used for soluble rCD4 expression⁷. A human IgG1 heavy-chain cDNA, obtained from a human spleen cDNA library using probes based on the published sequence⁴⁷, was inserted at a unique *Xba*I site found immediately 3' of the CD4 coding region in the same reading orientation as CD4. Synthetic 48-mer oligodeoxynucleotides, complementary to the 24 nucleotides at the borders of the desired CD4 and IgG1 fusion sites, were used as primers in the mutagenesis reactions using the plasmid described above as the template⁴⁸.

particularly interesting class of these CD4-immunoglobulin hybrids 'immunoadhesins', because they contain part of an adhesive molecule²⁰ linked to the immunoglobulin Fc effector domain.

Synthesis of CD4 immunoadhesins

CD4 is an integral membrane protein with an extracellular region comprising four domains with homology to immunoglobulin variable domains^{21,22} (Fig. 1). Soluble CD4 derivatives consisting of this extracellular region bind gp120 with the same affinity as cell-surface CD4 (ref. 7). CD4 variants containing only domains 1 and 2 also bind gp120^{17,18}, but the affinity of this interaction is not known. We constructed a series of hybrid molecules consisting of the first two or all four immunoglobulin-like domains of CD4 fused to the constant region of antibody heavy and light chains (Fig. 1).

We investigated the synthesis and secretion of these hybrids using transient expression in a human embryonic kidney-derived cell line. As shown in Fig. 2, immunoglobulin light and heavy

chains are efficiently expressed in these cells, and light chain is efficiently secreted, but heavy chain is not unless a light chain is coexpressed. Thus the rules governing immunoglobulin chain secretion in these cells are the same as those for plasma or other lymphoid cells²³. We first constructed hybrids that fused CD4 with the constant regions of murine κ - or γ_1 -chains. These hybrids contained either the first two or all four immunoglobulin-like domains of CD4, linked at a position chosen to mimic the spacing between disulphide-linked cysteines seen in immunoglobulins (Fig. 1). As expected, the CD4- κ hybrids were secreted well, whereas hybrids between CD4 and mouse γ_1 -chain were expressed but not secreted unless a κ -chain or a CD4- κ hybrid was present.

A different and unexpected picture emerged when analogous CD4-heavy-chain hybrids were constructed using the constant region of human IgG1 heavy chain instead of mouse heavy chain. Such hybrids, containing either the first two or all four immunoglobulin-like domains of CD4 (named $2\gamma_1$ and $4\gamma_1$ respectively), were secreted in the absence of wild-type or hybrid light chains (Fig. 2a). Both $2\gamma_1$ and $4\gamma_1$ could be directly immunoprecipitated using *Staphylococcus aureus* protein A, which binds the Fc portion of IgG1, indicating that the protein A-binding sites of these constructs are fully functional. Indeed, both molecules can be purified to near homogeneity on protein A columns (Fig. 2b).

Structure of CD4 immunoadhesins

We examined the subunit structure of these immunoadhesin molecules using SDS-polyacrylamide gels (Fig. 2b). Without any reducing agent, the apparent relative molecular mass (M_r) of each construct doubled, demonstrating that both immunoadhesins are disulphide-linked dimers. The hinge region of each immunoadhesin contains three cysteine residues, one normally involved in disulphide bonding to light chain, the other two in the intermolecular disulphide bonds between the two heavy chains in IgG. As the molecules are dimers at least one, and perhaps all three, of these cysteine residues are involved in intermolecular disulphide bonds. We examined the capacity of $2\gamma_1$ and $4\gamma_1$ to form disulphide links with light chains. When an immunoadhesin construct was cotransfected with a light chain, the light chain produced could be precipitated by protein A. Mutagenic substitution of the first hinge-region cysteine with alanine abolished light-chain bonding, but did not affect dimerization (data not shown), indicating that this cysteine bonds the light chain in these hybrids, as in normal IgG. Thus the disulphide bond structure of these immunoadhesins seems to be analogous to that of immunoglobulins.

gp120 binding

To determine whether our immunoadhesins retain the ability to bind gp120 with high affinity, and whether the first two immunoglobulin-like domains are sufficient, we carried out saturation binding analyses with radioiodinated gp120. Binding is saturable, showing a simple mass action curve (Fig. 3a). The dissociation constant (K_d) for the interaction of each immunoadhesin with gp120, calculated by Scatchard analysis (Fig. 3a, inset), was indistinguishable from that of soluble rCD4 ($\sim 10^{-9}$ M) (Table 1). Thus, the N-terminal 170 amino acids of CD4 are sufficient for high-affinity binding. As these immunoadhesins are homodimeric, they should each have two gp120-binding sites. We examined this possibility by coating plastic microtitre wells with gp120, then adding soluble CD4 or immunoadhesins. Both immunoadhesins could bind added labelled gp120, whereas soluble rCD4, with only one gp120 binding site, could not (J. Porter and S. C., unpublished results). To confirm the bivalent nature of $2\gamma_1$ and $4\gamma_1$, we examined their ability to agglutinate sheep red blood cells coated with gp120. Again, both CD4 immunoadhesins, but not soluble rCD4, agglutinated the cells, showing that binding to gp120 molecules on different cells is not sterically hindered.

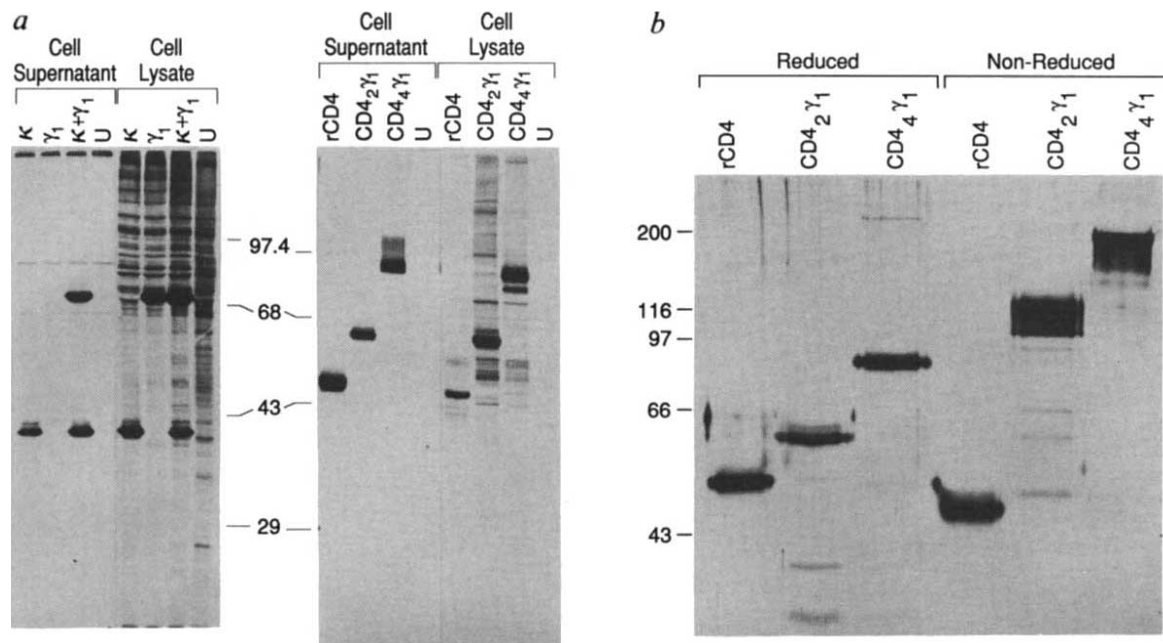


Fig. 2 Expression, secretion and subunit structure of CD4 immunoadhesins and soluble rCD4. *a*, Expression and secretion of mouse immunoglobulins, soluble rCD4 and CD4 immunoadhesins expressed in mammalian cells. Cells were transfected with vectors directing the expression of murine κ -light chain (lanes κ) or γ 1-heavy chain (lanes γ 1) individually or together (lanes $\kappa + \gamma$ 1), vectors encoding soluble rCD4 (lanes rCD4), and the CD4 immunoadhesins 2 γ 1 (lanes CD4₂ γ 1) or 4 γ 1 (lanes CD4₄ γ 1). After metabolic labelling with [³⁵S]methionine, cell supernatants and cell lysates were analysed by immunoprecipitation. Lanes U, untransfected cells. *b*, Subunit structure of secreted CD4 immunoadhesins and soluble rCD4. Soluble rCD4, 2 γ 1 and 4 γ 1 were purified from culture supernatants of transfected cells and analysed by electrophoresis on a 7.5% SDS-polyacrylamide gel. Samples were prepared in buffer with 10 mM dithiothreitol (DTT) (reducing conditions) or without DTT (non-reducing conditions). The positions of relative molecular mass standards are indicated (in thousands). Both immunoadhesins behaved as disulphide-linked dimers; in contrast, soluble rCD4 which is monomeric, displayed only a minor change in mobility upon reduction of its intra-molecular disulphide bonds.

Methods. *a*, Cells were transfected by a modification of the calcium phosphate procedure, labelled with [³⁵S]methionine, and cell lysates prepared as described⁴⁹. Immunoprecipitation analysis was carried out as previously described⁷, with the exception that no preadsorption with Pansorbin (Calbiochem) was done, and the precipitating antibodies used were 2 μ l of rabbit anti-mouse IgG serum (Cappel) for mouse IgG heavy and light chains, 0.25 μ g of OKT4A (Ortho) for soluble rCD4, and no added antibody (Pansorbin only) for the CD4 immunoadhesins. Immunoprecipitated proteins were resolved on 10% SDS-PAGE gels, and visualized by autoradiography. *b*, CD4 immunoadhesins were purified from transfected cell supernatants by protein A affinity chromatography followed by ammonium sulphate precipitation. Purified proteins were subjected to SDS-PAGE under both reducing and non-reducing conditions and visualized by silver staining.

In vivo plasma half-life

We examined whether the immunoadhesins share the long *in vivo* half-life of antibodies. Studies of rCD4 in rabbits provide clearance data that extrapolate well to other species, including humans (J.M., unpublished results). The change in plasma concentration with time for each of the three CD4 analogues in rabbits is shown in Fig. 4. Analysis of these data reveals that soluble rCD4 has a terminal half-life in rabbits of ~15 min, whereas 4 γ 1 and 2 γ 1 have terminal half-lives of ~7 and 48 h, respectively (Table 1). Thus the half-life of 2 γ 1 in rabbits is nearly 200 times longer than that of rCD4 and comparable to that of human IgG in rabbits (4.7 days)²⁴. The half-life of 2 γ 1 in humans is expected to be longer than that in rabbits, because of the decreased proportional blood flow to eliminating organs

as species increase in size²⁵, and should be comparable with that of human IgG1 (21 days).

Our results confirm our initial hypothesis that, as in the case of immunoglobulin itself, one can increase the stability of a rapidly cleared molecule (Fab or rCD4) by fusing it to a long-lived molecule, Fc. The swift clearance of rCD4 is probably largely due to its size, M_r 55,000, which means it is just small enough to be cleared efficiently by renal filtration. One component in the increased half-lives of these molecules is therefore probably their larger size; but this cannot be the whole story as 4 γ 1, although larger than 2 γ 1, has a shorter half-life. Both 4 γ 1 and rCD4, but not 2 γ 1, contain two CD4-derived Asn-linked carbohydrate sites which are glycosylated in rCD4 (R. Harris and M. Spellman, unpublished results); these sugar moieties

Table 1 Properties of CD4 immunoadhesins and soluble rCD4

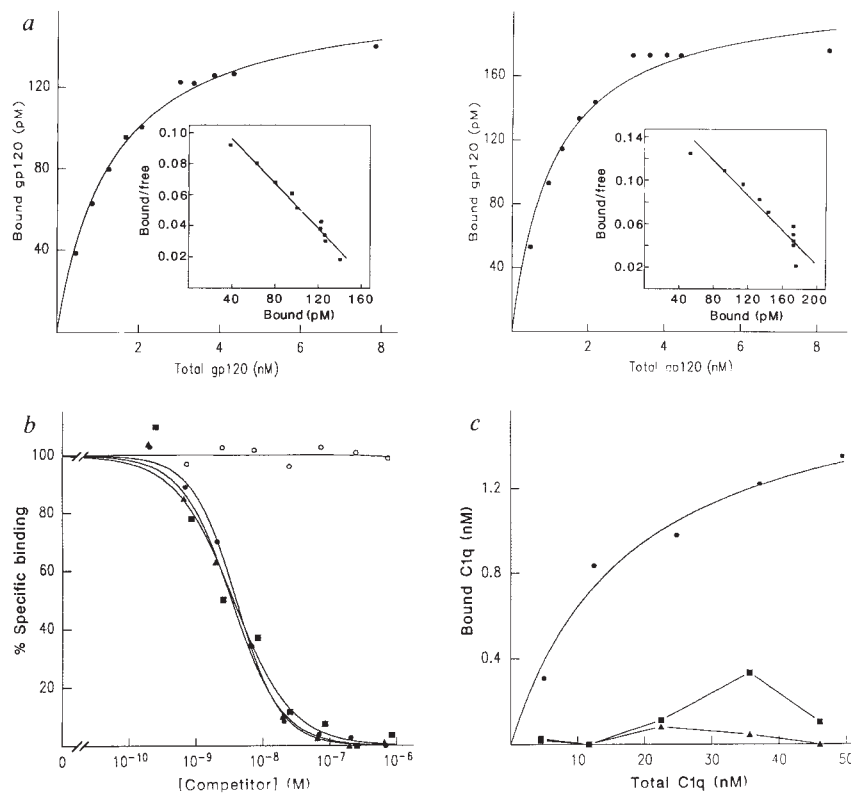
	Calculated M_r	Subunit structure	gp120 binding (nM)*	Blocks infectivity		Plasma half-life in rabbits (hours)†	Fc binding (nM)*	Complement binding	Protein A binding
				T cells	MØ				
rCD4	41,000	monomer	2.3 ± 0.4	Yes	Yes	0.25 ± 0.01	—	No	No
4 γ 1	154,000	dimer	1.2 ± 0.1	Yes	Yes	6.7 ± 1.1	2.3 ± 0.7	No	Yes
2 γ 1	112,000	dimer	1.4 ± 0.1	Yes	Yes	48.0 ± 8.6	2.6 ± 0.3	No	Yes
IgG1	146,000	tetramer (H ₂ L ₂)	—	—	—	113‡	3.2 ± 0.2	Yes	Yes

* Standard error of the mean was determined using the Inplot and Scatplot programs (see Fig. 3 legend). † Standard deviation indicated in hours.

‡ Determined in ref. 24 (IgG1 has a half-life of 21 days in humans).

Fig. 3 Binding properties of CD4 immunoadhesins. **a**, Gp120 saturation binding analysis of CD4 immunoadhesins. Immunoadhesin proteins 4 γ 1 (left) or 2 γ 1 (right) in transfected cell supernatants were incubated with increasing concentrations of purified soluble rgp120 (ref. 50) radioiodinated with lactoperoxidase. The lines drawn for the binding curves and for the Scatchard plots of the data (shown in the insets) represent the best fit as determined by unweighted least-squares linear regression analysis. Dissociation constants calculated from these results and from binding studies of gp120 to soluble rCD4 performed in parallel are given in Table 1. **b**, Binding of CD4 immunoadhesins to Fc γ receptors on U937 cells. Competition binding analysis was carried out by mixing 0.1 μ g ml $^{-1}$ of 125 I-labelled human IgG1 (Calbiochem) with increasing concentrations of purified human IgG1 (solid circle), 2 γ 1 (solid square), 4 γ 1 (solid triangle), or soluble rCD4 (open circle) proteins. Curves drawn represent the best fit as determined by unweighted least-squares nonlinear (IgG1, 2 γ 1 and 4 γ 1) or linear (rCD4) regression analysis. Dissociation constants calculated from these results are shown in Table 1. **c**, C1q saturation binding analysis of CD4 immunoadhesins. Purified anti-gp120 IgG2a mouse monoclonal antibody (solid circle), 2 γ 1 (solid square), or 4 γ 1 (solid triangle) proteins were aggregated by binding to gp120-coupled Sepharose, and incubated with increasing concentrations of purified human C1q (Calbiochem) radioiodinated with lactoperoxidase. The curve drawn for the anti-gp120 monoclonal antibody (mAb) represents the best fit as determined by least-squares nonlinear regression analysis; the dissociation constant for C1q binding to this gp120-aggregated anti-gp120 mAb was $\sim 1.8 \times 10^{-8}$ M.

Methods. **a**, Gp120 saturation binding analysis was carried out as described⁷ except that gp120-CD4 immunoadhesin complexes were collected directly onto Pansorbin: binding was comparable to that observed when complexes were collected with OKT4A as for soluble rCD4. Specifically bound 125 I-labelled gp120 was determined from the difference in binding in the presence or absence of a 1,000-fold excess of unlabelled rgp120 and is plotted against the total 125 I-labelled gp120 concentration. **b**, FcR binding analysis was done essentially as described²⁷ except that after centrifugation free IgG1 was removed by aspiration of the aqueous and oil layers. Mixtures of 125 I-labelled human IgG1 and IgG1, CD4 immunoadhesins or soluble rCD4 were incubated with U937 cells (2×10^6 cells per tube) for 60 min at 4°C. Specific binding was calculated by subtracting residual nonspecific binding (<25% of specific binding) which could not be competed out by a 1,000-fold excess of unlabelled human IgG1. **c**, C1q binding analysis was done essentially as described²⁹, except that gp120 coupled to CNBr-activated Sepharose 6B (Pharmacia) was used as the solid support to aggregate CD4 immunoadhesins or the anti-gp120 mouse mAb. Proteins were adsorbed to gp120 coupled-beads, incubated with varying concentrations of 125 I-labelled C1q, and bound and free C1q were then separated by centrifugation through 20% sucrose. Specific binding was determined from the difference in binding in the presence or absence of added antibody or immunoadhesin. All data analysis was carried out using the Inplot and Scatplot programs (R. Vandlen, Genentech). Scatplot was modified from the Ligand program (P. Muncy, NIH).



may facilitate clearance by receptors in the liver. The charge of the molecule may also be important, as the CD4 portion of 4 γ 1 contributes a net excess of eleven positively charged amino acids on 4 γ 1, but only three on 2 γ 1. This may increase uptake of rCD4 and 4 γ 1 onto anionic surfaces, accelerating their clearance from the circulation.

Fc receptor and complement binding

Two major mechanisms for the elimination of pathogens are mediated by the Fc portion of specific antibodies. Fc activates the classical pathway of complement, ultimately resulting in lysis of the pathogen, whereas binding to cell Fc receptors can lead to ingestion of the pathogen by phagocytes or lysis by killer cells. The binding sites for Fc cell receptors and for the initiating factor of the classical complement pathway, C1q, are found in the constant region of heavy chain²⁶ (the CH2 domain for C1q²⁷ and the region linking the hinge to CH2 for Fc cell receptors²⁸). We aimed to incorporate both of these functions into the immunoadhesins. We chose the IgG1 subtype to supply the Fc domain because IgG1 is the best compromise between Fc binding, C1q binding, and long half-life. We show below that the immunoadhesins bind FcR well, but do not bind C1q.

Three types of Fc cell receptors are known to be expressed on a variety of leukocytes. Of these FcRI, principally expressed

on mononuclear phagocytes, is the only one which binds monomeric human IgG1 with high affinity²⁶. We used competition binding analysis with FcRI receptors on the U937 monocyte/macrophage cell line to characterize the Fc receptor binding of 2 γ 1 and 4 γ 1. Direct saturation binding analysis with human IgG1 gave a K_d of $\sim 3 \times 10^{-9}$ M. In competition binding analyses, the two CD4 immunoadhesins, but not rCD4, bound to Fc receptors on U937 cells to the same extent and with an affinity indistinguishable from human IgG1 (Fig. 3b, Table 1).

We examined the ability of the immunoadhesins to bind to the first component of the classical pathway of complement, C1q, by saturation binding analysis. Because binding of C1q increases with the aggregation state of the antibody, with an affinity of $\sim 10^{-4}$ for monomers and $\sim 10^{-8}$ for tetramers of IgG²⁶, we first aggregated the immunoadhesin using gp120 linked to Sepharose. As a positive control, we measured C1q binding to an anti-gp120 mouse IgG2a monoclonal antibody, (which like human IgG1 binds C1q with high affinity²⁹) aggregated by the same gp120-Sepharose. The affinity of the mouse antibody for C1q determined by Scatchard analysis was 1.8×10^{-8} M (Fig. 3c), comparable to that observed for other mouse IgG2a and for human IgG1 antibodies. In contrast, neither immunoadhesin bound C1q to any detectable extent (Fig. 3c),

Fig. 4 Pharmacokinetics of CD4 immunoadhesins and soluble rCD4. Shown are the mean plasma concentrations (ng ml^{-1}) for 2 γ 1 (triangles), 4 γ 1 (squares), and rCD4 (circles) following a single intravenous administration in rabbits. *a*, Time course of plasma clearance over the first 120 minutes; *b*, time course over 8 days after injection of the CD4 analogues.

Methods. Ten female New Zealand white rabbits (Rabbitek, Modesto, California) were injected intravenously (via an ear vein catheter) with a single bolus dose ($40 \mu\text{g kg}^{-1}$ in a volume of 1 ml) of either rCD4 ($n=2$), 4 γ 1 ($n=4$), or 2 γ 1 ($n=3$). Blood samples were obtained from an arterial catheter in the opposite ear; after 24 hours, blood samples were obtained by venipuncture. Plasma concentrations of each protein were determined by an enzyme-linked immunosorbent assay. This capture assay used two antibodies, including an anti-CD4 monoclonal directed against the gp120-binding site (and capable of blocking gp120 binding), and thus provided a sensitive assay for CD4-containing molecules that are still capable of binding gp120. Exponential equations were fitted to the data of individual rabbits using a nonlinear least squares regression program NONLIN84[®] (Statistical Consultants, Lexington, Kentucky). The concentration (C , ng ml^{-1}) versus time (t) data for rCD4 were best described by a biexponential equation $C = 541 e^{-1.34t} + 620 e^{-0.0472t}$, where time is in minutes; the average terminal half-life was 14.7 min, and the average clearance was $3 \text{ ml min}^{-1} \text{ kg}^{-1}$. The 4 γ 1 data were best described by a triexponential equation, $C = 546 e^{-211t} + 193 e^{-20.5t} + 46.8 e^{-2.54t}$, where time is in hours. The average terminal half-life was 6.7 hours, and the average clearance was $0.91 \text{ ml min}^{-1} \text{ kg}^{-1}$. The 2 γ 1 data were best described by a triexponential equation, $C = 153 e^{-53.2t} + 342 e^{-2.19t} + 183 e^{-0.351t}$, where time is in hours. The average terminal half-life was 48 hours, and the average clearance was $0.039 \text{ ml min}^{-1} \text{ kg}^{-1}$.

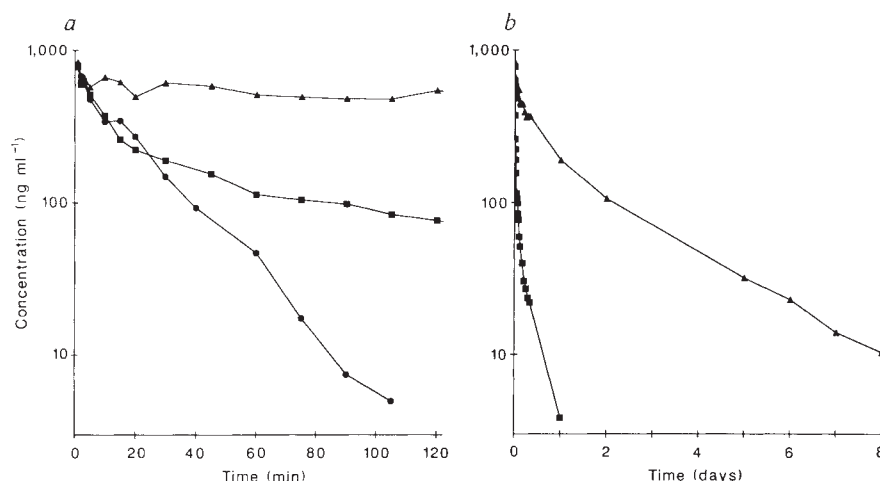
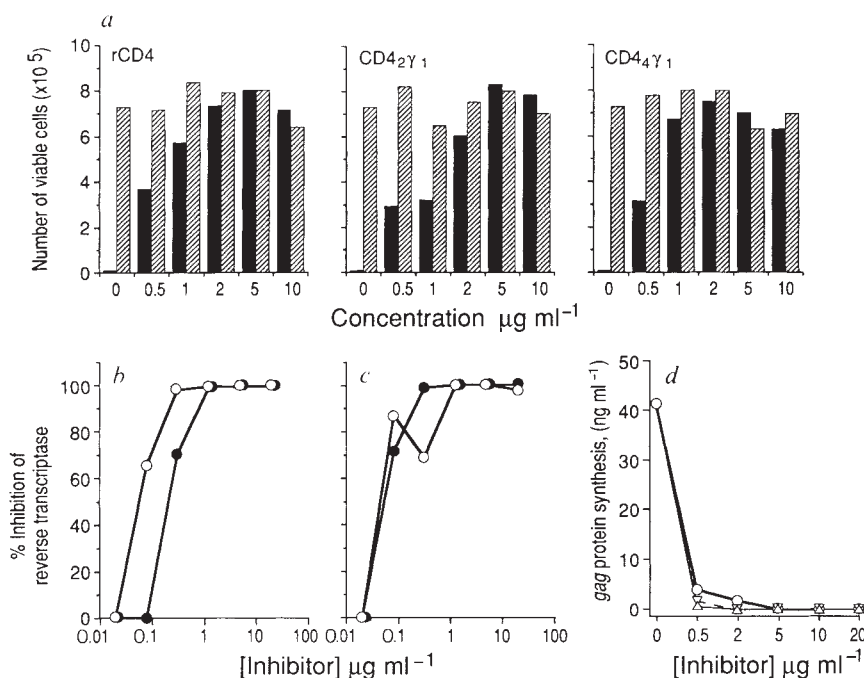


Fig. 5 Inhibition of HIV-1 infectivity by CD4 immunoadhesins and soluble rCD4. *a*, Inhibition of the cytopathic effects on ATH8 cells by HIV-1 was examined as described³² with the HTLV-IIIB isolate³¹. The number of viable cells at day 10 after infection is shown for varying concentrations of each molecule in the presence (solid bars) or absence (shaded bars) of added virus. The absence of an effect of each CD4 analogue on cell number in the absence of virus indicates that none of these molecules inhibited cell growth. *b*, Inhibition of infection of H9 cells by HIV-1 was carried out as described⁷ with the HTLV-IIIB isolate. Reverse transcriptase activity was determined 7 days after infection and is given as the percentage of the level seen in the absence of inhibitor. Solid and open circles represent 2 γ 1 and 4 γ 1, respectively. *c*, Inhibition of infection of U937 cells by HIV-1 (HTLV-IIIB isolate) was carried out as described above for H9 cells. *d*, Inhibition of infection of fresh human monocytes by the monocytotropic HIV-1 isolate Ba-L (ref. 35). HIV-1 replication was determined by measuring the level of p24 gag antigen synthesis 10 days after infection using a commercial assay kit (Dupont). Circles, inverted triangles and triangles represent inhibition of p24 synthesis by soluble rCD4, 2 γ 1 and 4 γ 1, respectively.



although both did bind the gp120-Sepharose matrix in amounts comparable to the control antibody.

Thus, our immunoadhesins bind well to Fc receptors. It is perhaps surprising that they do not bind C1q. As far as is known, all the critical contact residues for C1q binding reside in the CH2 domain of the heavy chain²⁶, and are conserved among all the human IgG isotypes. However, these have varying abilities to mediate complement fixation. Thus steric hindrance or other aspects of protein conformation (for example, the segmental flexibility of antibodies³⁰) may be important.

Infectivity studies

Two systems were used to study the *in vitro* ability of CD4 immunoadhesins to block infection of CD4-bearing T cells by

the HIV-1 T-lymphotrophic isolate HTLV-IIIB (ref. 31). Infection with HIV-1 exerts a profound cytopathic effect on the human T-cell clone ATH8, with more than 98% of the cells being killed by day 10 after infection³² (Fig. 5a). Both CD4 immunoadhesins blocked cell killing with the same potency as soluble rCD4, without inhibiting cell proliferation; each CD4 analogue completely abolished cell killing at a concentration of $\sim 0.05 \mu\text{M}$ (Fig. 5a). Complete protection was also observed at comparable concentrations with a different HIV-1 isolate, HTLV-III RF, which is not neutralized by sera from animals immunized with rgp120 from the IIIB isolate⁵. We also examined the production of HIV-1 reverse transcriptase activity after infection of the H9 human T-cell line. Again, both immunoadhesins completely blocked virus production by day 7 (Fig. 5b), at

concentrations comparable to rCD4 (data not shown); moreover the potency of each CD4 analogue was markedly higher (~fivefold) than that observed in the ATH8 assay.

Monocyte infection

Because it has been suggested that antibodies present in sera from HIV-1 infected individuals may enhance the infectivity of HIV-1 in Fc receptor (FcR)-bearing cells such as primary blood monocytes³³, and monocyte cell lines³⁴, we examined the effect of rCD4 and CD4 immunoadhesins on HIV-1 infection of FcR-expressing cells of monocyte/macrophage origin. Both CD4 immunoadhesins completely blocked HIV-1 IIIB virus production in U937 cells at similar concentrations to those found to be effective on H9 cells (Fig. 5c), with a potency comparable to that of soluble rCD4 (data not shown). In another system, the replication of a monocytopathic HIV-1 isolate, Ba-L³⁵, in fresh monocytes was monitored by the production of p24 antigen. Soluble rCD4 completely blocked infection, indicating that infection of monocytes by the Ba-L isolate does involve CD4. Both CD4 immunoadhesins also completely blocked p24 production, at concentrations equal to or lower than rCD4 (Fig. 5d). Thus the CD4 immunoadhesins are at least comparable to soluble rCD4 in their ability to prevent infection of monocyte/macrophages by HIV; no evidence was found for enhancement of infection by immunoadhesins (or by soluble rCD4) in cells which express high affinity Fc receptors.

Implications for treatment of HIV-1 disease

Because the hallmark of HIV-1 disease is the specific destruction of CD4⁺ T cells, and the progression of infected individuals to AIDS closely parallels their decline in CD4⁺ T-cell number³⁶, it is reasonable to believe that the interaction of gp120 with CD4, either by direct HIV-1 infection of CD4⁺ cells or otherwise, underlies the killing of CD4⁺ cells. Therefore, if this interaction can be stopped it may be possible to prevent disease progression. But despite the logic of this hypothesis, the observation that only a very few lymphocytes are actively infected with HIV-1 *in vivo*³⁷ has posed a problem to those attempting to explain the causative role of HIV-1 in the aetiology of AIDS³⁸. Two observations may explain the 'catalytic' ability of HIV-1 to deplete CD4⁺ lymphocytes: first, a single infected cell can fuse many uninfected CD4⁺ cells to itself, creating an inviable mass^{39,40}, and second, gp120 is shed from the surface of HIV-1-infected cells and virions⁴¹, as its link to gp41, its anchor protein partner, is probably non-covalent. This shed gp120 binds to surface CD4 on uninfected cells with high affinity, and can result in their functional alteration^{42,43} or death by one of two pathways shown to operate *in vitro*. Bystander cells coated with gp120 bound to their CD4 surface molecules become targets for anti-gp120 antibodies produced by HIV-1 infected individuals and can be killed via antibody-dependent cell-mediated cytotoxicity⁴⁴. Also, MHC class II-positive CD4⁺ T cells can internalize gp120 bound tightly to CD4 on their surface, process it, and present peptides derived from it on their class II molecules, thus becoming sensitive, even at low gp120 concentrations, to lysis by gp120-specific cytotoxic T cells^{19,45,46}. The important common factor in all these proposed mechanisms of cell destruction is that gp120 must bind specifically to cell-surface CD4. If these mechanisms are important *in vivo*, this would imply that soluble rCD4 could intervene.

But to affect the disease noticeably, one would expect to need to maintain a high concentration of rCD4, which is hampered by its rapid clearance. Our approach to this problem was to fuse the gp120-binding domain of CD4 to a molecule well designed to avoid the clearance mechanisms of the body. Indeed, the Fc domain and CD4 sequences are structurally compatible, as the hybrid molecules have important properties of both parents. Thus, they bind gp120 and block infection of T cells by T-lymphotrophic HIV-1 and of monocytes by monocytopathic HIV-1. They are also comparable to antibodies in

their long plasma half-life and their ability to bind Fc receptors and protein A. This combination of properties allows both a better passive defence, due to the higher plasma concentrations attainable even with infrequent injection, and the possibility of actively attacking HIV-1 and infected cells. A high steady-state level also makes it more likely that effective concentrations will be attained in lymph and lymphatic organs, where HIV may be most active.

The high-affinity binding of the immunoadhesins to Fc receptors implies that mechanisms of pathogen elimination, such as phagocytic engulfment and killing by antibody-dependent cell-mediated cytotoxicity, may be recruited by these immunoadhesins to kill HIV-1 infected cells and virus. As it is possible that antibody-dependent cell-mediated cytotoxicity in an infected individual may be more a mechanism of pathology in HIV-1 infection than a protective response⁴⁴, it is important to note a difference between CD4 immunoadhesins and the patients' own anti-gp120 antibodies: the immunoadhesin, in contrast to antibody, cannot recognize gp120 bound to an uninfected CD4⁺ bystander cell, as gp120 has only a single binding site for CD4. Because placental transfer of antibody, unique to the IgG subclass, also proceeds through an FcR-dependent mechanism, CD4 immunoadhesins may also be transferred *in utero*. This may have implications for the prevention of perinatally transmitted HIV-1 infection.

Although it is not yet clear which of the functions of immunoglobulins will be advantageous when applied to HIV infection, we have taken the approach of trying to add all possible functions to our immunoadhesins. Once the structural requirements for the optimal molecule are established, functions can be tailored at will, as the parent antibody molecule is so well understood.

We thank Drs Paula Jardieu, Avi Ashkenazi and Stephen Sherwin for advice and helpful discussions, Dr Rebecca Ward for helpful discussions and critical reading of the manuscript, Steven Frie for performing CD4 enzyme-linked immunosorbent assays, Vivek Bajaj and Wally Tanaka for large scale cell culture, Drs R. Harris, M. Spellman and J. Porter for allowing us to cite unpublished data, Steve Williams for murine immunoglobulin cDNAs, Mark Vasser, Parkash Jhurani and Peter Ng for synthetic DNA, Dr Brian Fendly and Kim Rosenthal for anti-gp120 monoclonals, and Carol Morita and Kerrie Andow for preparation of the figures. R.B. and J.G. are supported by grants from the NIH and the US Defense Department.

Received 3 November; accepted 16 December 1988.

- Curran, J. *et al.* *Science* **239**, 610-616 (1988).
- Hu, S.-L. *et al.* *Nature* **328**, 721-723 (1987).
- Berman, P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 5200-5204 (1988).
- Prince, A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 6944-6948 (1988).
- Weiss, R. *et al.* *Nature* **324**, 572-575 (1986).
- Mitsuya, H. & Broder, S. *Nature* **325**, 773-778 (1987).
- Smith, D. *et al.* *Science* **328**, 1704-1707 (1987).
- Sattentau, Q. & Weiss, R. *Cell* **52**, 631-633 (1988).
- Janeway, C. *Nature* **335**, 208-210 (1988).
- Dalglish, A. *et al.* *Nature* **312**, 763-767 (1984).
- Klatzmann, D. *et al.* *Nature* **312**, 767-768 (1984).
- Maddon, P. *et al.* *Cell* **47**, 333-348 (1986).
- McDougal, J. *et al.* *Science* **231**, 382-385 (1986).
- Fisher, R. *et al.* *Nature* **331**, 76-78 (1988).
- Hussey, R. *et al.* *Nature* **331**, 78-81 (1988).
- Deen, K. *et al.* *Nature* **331**, 82-84 (1988).
- Trautnecker, A., Luke, W. & Karjalainen, K. *Nature* **331**, 84-86 (1988).
- Berger, E., Fuerst, T. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2357-2361 (1988).
- Siliciano, R. *et al.* *Cell* **54**, 561-575 (1988).
- Doyle, C. & Strominger, J. *Nature* **330**, 256-259 (1987).
- Maddon, P. *et al.* *Cell* **42**, 93-104 (1985).
- Clark, S., Jefferies, W., Barclay, A., Gagnon, J. & Williams, A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1649-1653 (1987).
- Dorai, H. & Moore, G. *J. Immun.* **139**, 4232-4241 (1987).
- Nakamura, R., Spiegelberg, H., Lee, S. & Wiegler, W. *J. Immun.* **100**, 376-383 (1968).
- Mordenti, J. *J. Pharmaceut. Sci.* **75**, 1028-1040 (1986).
- Burton, D. *Molec. Immun.* **22**, 161-206 (1985).
- Duncan, A. & Winter, G. *Nature* **332**, 738-740 (1988).
- Duncan, A., Woof, J., Partridge, L., Burton, D. & Winter, G. *Nature* **332**, 563-564 (1988).
- Leatherbarrow, R. & Dwek, R. *Molec. Immun.* **21**, 321-327 (1984).
- Feinstein, A., Richardson, N. & Taussig, M. *J. Immun. Today* **7**, 169-174 (1986).
- Gallo, R. *et al.* *Science* **224**, 500-505 (1984).
- Mitsuya, H. & Broder, S. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1911-1915 (1986).

33. Homsy, J., Tateno, M. & Levy, J. *Lancet* **i**, 1285-1286 (1988).
34. Takeda, A., Tuazon, C. & Ennis, F. *Science* **242**, 580-583 (1988).
35. Gartner, S. *et al. Science* **233**, 215-219 (1986).
36. Lane, H. & Fauci, A. *Rev. Immun.* **3**, 477-500 (1985).
37. Harper, M., Marselle, L., Gallo, R. & Wong-Staal, F. *Proc. natn. Acad. Sci. U.S.A.* **83**, 772-776 (1986).
38. Duesberg, P. *Science* **241**, 514 (1988).
39. Lifson, J., Reyes, G., McGrath, M., Stein, B. & Engelman, E. *Science* **232**, 1123-1127 (1986).
40. Sodroski, J., Goh, W., Rosen, C., Campbell, K. & Haseltine, W. *Nature* **322**, 470-474 (1986).
41. Schneider, I., Kaaden, O., Copeland, T. D., Orosian, S. & Hunsmann, G. *J. gen Virol.* **67**, 2533-2539 (1986).

42. Lynette, G., Hartzman, R., Ledbetter, J. & June, C. *Science* **241**, 573-576 (1988).
43. Kornfeld, H., Cruikshank, W., Pyle, S., Berman, J. & Center, D. *Nature* **335**, 445-448 (1988).
44. Iyerly, H., Matthews, T., Langlois, A., Bolognesi, D. & Weinhold, K. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4601-4605 (1987).
45. Lanzavecchia, A., Roosneck, E., Gregory, T., Berman, P. & Abrignani, S. *Nature* **334**, 530-532 (1988).
46. Sethi, K., Naher, H. & Stroehmann, I. *Nature* **335**, 178-181 (1988).
47. Ellison, J. P., Berson, B. J. & Hood, L. E. *Nucleic Acids Res.* **10**, 4071-4079 (1982).
48. Zoller, M. & Smith, M. *Nucleic Acids Res.* **10**, 6487-6500 (1982).
49. Muesing, M., Smith, D. & Capon, D. *Cell* **48**, 691-701 (1987).
50. Lasky, L. *et al. Cell* **50**, 975-985 (1987).

LETTERS TO NATURE

A 110-ms pulsar, with negative period derivative, in the globular cluster M15

A. Wolszczan*, S. R. Kulkarni†, J. Middleditch‡, D. C. Backer§, A. S. Fruchter|| & R. J. Dewey¶#

* Arecibo Observatory, Arecibo, Puerto Rico 00613

† Department of Astronomy, California Institute of Technology, Pasadena, California 91125, USA

‡ Computing and Communications Division, Los Alamos National Laboratory, New Mexico 87545, USA

§ Astronomy Department, University of California, Berkeley, California 94720, USA

|| Joseph Henry Laboratories and Physics Department, Princeton University, Princeton, New Jersey 08544, USA

¶ Center for Radiophysics and Space Research, Cornell University, Ithaca, New York, 14853, USA

We report the discovery of a 110-ms pulsar, PSR2127+11, in the globular cluster M15 (NGC7078)¹. The results of nine months of timing measurements place the new pulsar about 2" from the centre of the cluster, and indicate that it is not a member of a close binary system. The measured negative value of the period derivative, $\dot{P} \approx -2 \times 10^{-17} \text{ s s}^{-1}$, is probably the result of the pulsar being bodily accelerated in our direction by the gravitational field of the collapsed core of M15. This apparently overwhelms a positive contribution to $\dot{P}$ due to magnetic braking. Although PSR2127+11 has an unexpectedly long period, we argue that it belongs to the class of 'recycled' pulsars, which have been spun up by accretion in a binary system. The subsequent loss of the pulsar's companion is probably due to disruption of the system by close encounters with other stars^{2,3}.

The discoveries of millisecond pulsars in globular clusters M28 (ref. 4) and M4 (ref. 5) led us to survey all clusters accessible to the 305-m Arecibo radio telescope ($0^\circ \leq \delta \leq 38^\circ$). A dual-polarization, 40-MHz-bandwidth signal at 1415 MHz was passed through the Arecibo digital correlator, sampled with 128 lags every 506.6 μs , and recorded on tape. The relatively high central radio frequency ensured an almost interference free signal and minimized the effects of interstellar dispersion and scattering, which can be significant for distant, low-galactic-latitude clusters. M15 was observed on 28 December 1987 for 90 minutes, which corresponds to ~ 11 million samples.

The data were analysed at both the Cornell National Supercomputer Facility (IBM 3090-600E) and the Los Alamos National Laboratory (Cray X-MP). Both analyses involved preliminary dedispersion of the multichannel data at 128 or 64 trial dispersion measures, followed first by one-dimensional Fourier transformation of the dedispersed time series and then by a search for harmonically related spikes in the resultant power spectra. The Cray X-MP analysis used the full, 11-million-sample data arrays to obtain maximum sensitivity with regard to isolated pulsars. The data analysed with the IBM supercom-

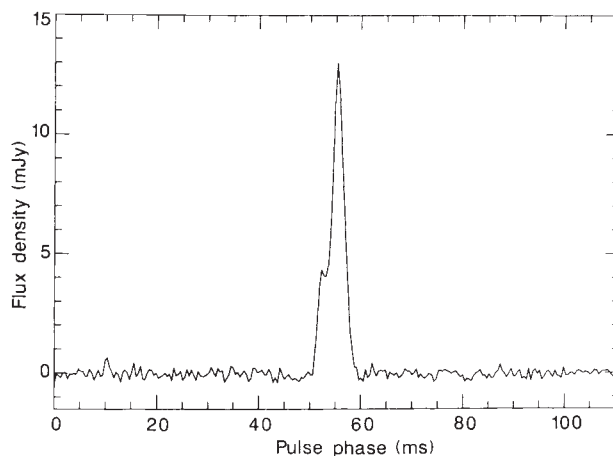


Fig. 1 The average pulse profile of PSR2127+11 at 1415 MHz. The effective resolution is $\sim 800 \mu\text{s}$ and the integration time is 7 hours.

puter were divided into five 2-million-sample blocks, which were treated separately to maintain high sensitivity to binary pulsars with short orbital periods. The nominal 6σ sensitivities of these two analysis schemes were 0.05 mJy and 0.1 mJy respectively, for the periods down to ~ 2.5 ms.

The data analysis at Cornell revealed the presence of a 110-ms, high- Q periodicity in the received signal with dispersion measure $DM \approx 60 \text{ pc cm}^{-3}$. This detection was subsequently confirmed at Los Alamos. Further observations made at Arecibo on 20 and 21 February 1988 confirmed the discovery of a 110-ms pulsar. The average pulse profile of PSR2127+11 observed at 1415 MHz is shown in Fig. 1. The pulsar parameters, derived from our twice-weekly timing observations over nine months, are summarized in Table 1. Errors quoted are the standard 3σ errors of a model fit to the observed pulse arrival times.

Although the precise timing and Very Large Array (VLA) positions of PSR2127+11 will become known soon, the present positional accuracy is sufficient to conclude that the pulsar is located well within the $6''$ core radius of the cluster, $2.0''$ west and $0.6''$ north of the centre⁶. The dispersion measure of PSR2127+11, $DM = 67.25 \text{ pc cm}^{-3}$, agrees well with that expected from a simple model of the galactic electron density distribution⁷, given the distance, $D = 9.7 \text{ kpc}$, and galactic coordinates,

Table 1 Measured parameters of the pulsar PSR2127+11

Pulsar period	$0.11066470954 \pm 0.0000000001 \text{ s}$
Period derivative	$(-20 \pm 1) \times 10^{-18} \text{ s s}^{-1}$
Epoch	JD 2447213.15
Dispersion measure	$67.25 \pm 0.05 \text{ pc cm}^{-3}$
Flux density (430 MHz)	$1.7 \pm 0.4 \text{ mJy}$
Flux density (1400 MHz)	$0.2 \pm 0.05 \text{ mJy}$
Right Ascension (B1950.0)	$21^{\text{h}}27^{\text{m}}33.22^{\text{s}} \pm 0.01$
Declination (B1950.0)	$11^{\circ}56'49.4'' \pm 0.3$
Distance	9.7 kpc

Present address: Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91125, USA.

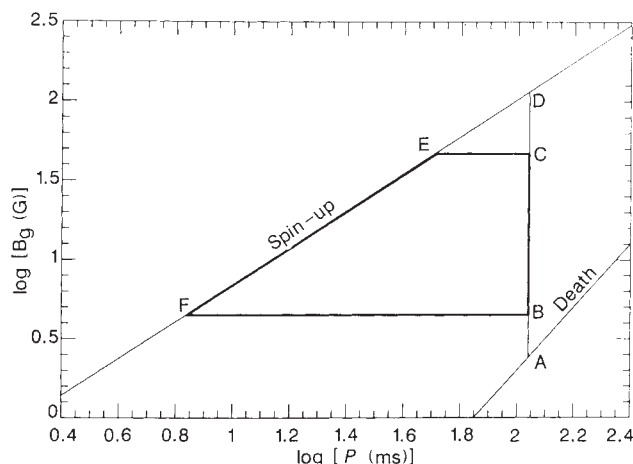


Fig. 2 A log-log plot of magnetic field B (scaled by 10^9 G) against period P for pulsars, showing the evolutionary constraints on P_i and B_0 for PSR2127+11. The pulsar evolves horizontally from the 'spin-up' line to line AD, which marks the present period. The segments BF and CE delimit the range of possible evolutionary tracks.

$l = 65^\circ$, $b = 27^\circ$, of M15. We conclude that PSR2127+11 is indeed associated with M15 and is at a projected distance from the centre of $2.1 \pm 0.5''$ (combined error).

As the pulsar is situated $1.8'' \pm 0.4''$ south of the optical position of the X-ray binary X2127+119 (AC 211) (ref. 8), the coincidence of these two objects is unlikely. Our timing observations confirm that PSR2127+11 is not associated with this 8.5-hour binary system.

All radio pulsars are believed to be rotationally powered and to be spinning down as a result of electromagnetic torques. This is consistent with their observed positive period derivatives, $\dot{P}$. The unusual negative $\dot{P}$ of PSR2127+11 can be explained if the pulsar is accelerated by $a_0 = c\dot{P}/P = 5 \times 10^{-6} \text{ cm s}^{-2}$. The true value of the acceleration must be higher than a_0 because timing measurements are sensitive only to the radial component of pulsar velocity. Moreover, the measured $\dot{P}$ used to derive a_0 includes an unknown, positive contribution from the pulsar's magnetic braking.

The observed acceleration is the result of combined effects of the mean gravitational potential of the cluster core and the potential of nearby stars⁹. Assuming the central stellar density of the cluster estimated in refs 10 and 11, the mean potential contributes an acceleration of $\sim 10^{-6} \text{ cm s}^{-2}$, which is significantly less than a_0 . A close encounter of the pulsar with a nearby field star (impact parameter ≤ 250 AU) could account for the value of a_0 but such events are very rare^{2,3}.

Clearly, a measurement of $\dot{P}$ will be helpful in further studies of the acceleration of PSR2127+11 (ref. 9). If the acceleration is caused by the pulsar's interaction with surrounding stars, then using a crude estimate of the mean timescale for a change in acceleration, $t \approx 1,000$ yr, we obtain $\dot{P}/P \approx a/ct = 5 \times 10^{-27} \text{ s}^{-2}$. An effect of this magnitude should be measurable to $\sim 5\%$ accuracy in about 2 years, assuming 10 pulse-arrival-time measurements per year with our present $\sim 50\text{-}\mu\text{s}$ r.m.s. timing residual¹². If acceleration is caused by the potential of the cluster core, the timescale for change is several orders of magnitude longer and therefore the contribution to $\dot{P}$ would be undetectable.

PSR2127+11 probably belongs to the class of 'recycled' pulsars which includes other globular-cluster pulsars and the galactic-disk binary and millisecond pulsars. Recycled pulsars are neutron stars that are first slowed by magnetic braking and then spun up by accretion from their binary companions. When the companion reaches the end of its stellar evolution, accretion

stops and the neutron star becomes a radio pulsar. The initial spin period, P_i , of the recycled pulsar depends on its magnetic field strength¹³: $P_i \approx 1.9 B_0^{6/7} \text{ ms}$, where B_0 is in units of 10^9 gauss. This relationship defines the so-called 'spin-up' line in the B/P diagram (Fig. 2). It is now believed that the magnetic fields of recycled neutron stars do not decay^{14,15}; the evolution of recycled pulsars proceeds horizontally until they reach the 'death line', defined by $B_0/P^2 \approx 2 \times 10^{-4}$, that is predicted by some models of pulsar emission¹⁶. Consequently, PSR2127+11 was born at some point on the 'spin-up line' and is now located somewhere along the segment AD in Fig. 2. Because $B_0 \approx (\dot{P}P)^{1/2}$ and $\dot{P}$ are corrupted as discussed above, the evolution of PSR2127+11 cannot be satisfactorily constrained. Purely formal upper limits to P_i and B_0 are defined by the probable condition that $\dot{P} < 2 \times 10^{-17} \text{ s s}^{-1}$, which gives $B_0 < 47$ and $P_i < 52 \text{ ms}$ (segment CE). The lower limits to P_i and B_0 can be estimated by requiring that the magnetic-field-dependent evolution of the pulsar until the present epoch cannot take longer than the Hubble time, $\sim 1.5 \times 10^{10} \text{ yr}$. This leads to lower limits $P_i \geq 7 \text{ ms}$ and $B_0 \geq 4.5$, as shown by segment BF in Fig. 2.

PSR2127+11, like PSR1821-24 in the globular cluster M28 (ref. 17), is not a member of a close binary system. Thus, it has somehow managed to dispose of its companion after or during the spin-up episode. There are essentially two mechanisms through which this can happen: by disruption of a binary through stellar encounters^{2,3,18} and by ablation of the companion¹⁹⁻²¹, a suggestion that resulted from the recent discovery of PSR1957+20 (ref. 22). The latter mechanism requires that P_i of the pulsar must have been no more than a few milliseconds, because it is the stored rotational energy that ablates the companion. Our lower limit of $P_i \geq 7 \text{ ms}$ for PSR2127+11 therefore makes the ablation model unlikely. A smaller value of P_i would be possible if the true spin-up line in Fig. 2 were displaced to the left of its usual position^{17,23}. However, all but one of the recycled pulsars lie to the right of the canonical spin-up line; the only exception is PSR1821-24, for which, like PSR2127+11, $\dot{P}$ may be dominated by gravitational effects. Thus we feel that the ablation mechanism, requiring a redefinition of the rebirth line, is unlikely. On the other hand, the high stellar density in the core of M15, implied by the large acceleration that is inferred from our timing observations, suggests that dynamical processes (encounters with other stars) could have disrupted the binary system.

The radio luminosity of PSR2127+11, $L \equiv S_{430} d^2 = 170 \text{ mJy kpc}^2$ (S_{430} is the pulsar's flux density at 430 MHz), is considerably greater than the minimum luminosity of $\sim 1 \text{ mJy kpc}^2$ (ref. 24), so that the pulsar may be just one of many such objects. M15 itself has at least one other neutron-star system, X2127+119, and it is entirely possible that there are more pulsars in M15 and in other clusters, although most of them will remain undetected because globular clusters are very distant.

Arecibo Observatory is part of the National Astronomy and Ionosphere Center, which is operated by Cornell University under contract with the NSF.

Received 19 September; accepted 15 December 1988.

1. Wolszczan, A., Middleditch, J., Kulkarni, S. R., Backer, D. C. & Fruchter, A. S. *IAU Telegram* No. 4552 (1988).
2. Romani, R. W., Kulkarni, S. R. & Blandford, R. D. *Nature* **329**, 309-310 (1987).
3. Verbunt, F., van den Heuvel, E. P. J., van Paradijs, J. & Rappaport, S. A. *Nature* **329**, 312-314 (1987).
4. Lyne, A. G., Brinklow, A., Middleditch, J., Kulkarni, S. R., Backer, D. C. & Clifton, T. R. *Nature* **328**, 399-401 (1987).
5. Lyne, A. G., Biggs, J. D., Brinklow, A., Ashworth, M. & McKenna, J. *Nature* **332**, 45-47 (1988).
6. Shawl, S. J. & White, R. F. *Astr. J.* **91**, 312-316 (1986).
7. Lyne, A. G., Manchester, R. N. & Taylor, J. H. *Mon. Not. R. astr. Soc.* **213**, 613-639 (1985).
8. Auriere, M., Geffert, M. & Koch-Miramond, L. *IAU Telegram* No. 4559 (1988).
9. Blandford, R. D., Romani, R. W. & Applegate, J. H. *Mon. Not. R. astr. Soc.* **225**, 51P-53P (1987).
10. Peterson, C. J. & King, I. R. *Astr. J.* **80**, 427-437 (1975).
11. Webbink, R. F. in *IAU Symp. 113* (eds Goodman, J. & Hut, P.) 541-577 (Reidel, Boston, 1985).
12. Blandford, R. D., Narayan, R. & Romani, R. W. *J. Astrophys. Astr.* **5**, 369-388 (1984).
13. Alpar, M. A., Cheng, A. R., Ruderman, M. A. & Shaham, J. *Nature* **300**, 728-731 (1982).

14. Kulkarni, S. R. *Astrophys. J.* **306**, L85–L89 (1986).
15. Van den Heuvel, E. P. J. & van Paradijs, J. A. *Nature* **322**, 153–155 (1986).
16. Manchester, R. N. & Taylor, J. H. *Pulsars* (Freeman, San Francisco, 1977).
17. Foster, R. S., Backer, D. C., Taylor, J. H. & Goss, W. M. *Astrophys. J.* **326**, L13–L15 (1988).
18. Verbunt, F. & Rappaport, S. A. *Astrophys. J.* **332**, 193–198 (1988).
19. Phinney, E. S., Evans, C. R., Blandford, R. D. & Kulkarni, S. R. *Nature* **333**, 832–834 (1988).
20. Kluzniak, W., Ruderman, M. A., Shaham, J. & Tavani, M. *Nature* **334**, 225–227 (1988).
21. Van den Heuvel, E. P. J. & van Paradijs, J. A. *Nature* **334**, 227–228 (1988).
22. Fruchter, A. S., Stinebring, D. R. & Taylor, J. H. *Nature* **333**, 237–239 (1988).
23. White, N. E. & Stella, L. *Nature* **332**, 416–418 (1988).
24. Dewey, R. J. thesis, Princeton University (1984).

Old cold dust heated by supernova 1987A

P. F. Roche*, D. K. Aitken†‡, C. H. Smith†
& S. D. James†

* Royal Observatory Edinburgh, Blackford Hill, Edinburgh
EH9 3HJ, UK

† Department of Physics, University College, ADFA, University of
New South Wales, Campbell, ACT 2600, Australia

The mid-infrared emission (at wavelengths near 10 μm) from SN1987A, which had been fading steadily since reaching a peak about 120 days after outburst, began to increase again on about day 450¹. This increase has continued up to at least day 578, and is probably due to heating of dust grains by light emitted at the optical maximum. Here we present additional observations and discuss the mass, location and nature of the emitting dust. The dust grains cannot have formed in the ejecta of the supernova, nor do they appear to be symmetrically distributed about it; rather, they are located at a distance of about 1 light year from, and mostly behind, the supernova.

The observations described here were obtained with the UCL array spectrometer on the Anglo-Australian telescope on days 465, 517 and 578 after the explosion of SN1987A on 23 February 1987; Fig. 1a shows the 8–13 μm spectra obtained on days 465 and 578. These were obtained through a 5.6-arcsec-diameter circular aperture with a spectral resolution of 0.09 μm . Flux calibration and correction for atmospheric absorption were made with respect to the bright star α Car. In addition, spatial scans across the supernova and nearby bright stars, to search for extended emission, were made on day 580 through a 2-arcsec aperture. Unfortunately, direct comparison of the spatial extent of the emission in the continuum with the emission lines in the supernova was not possible because, owing to instrumental problems, the spatial observations covered the spectrum only between 8 and 10 μm . Gaussian fits to east–west and north–south scans across the supernova yield FWHM (full width at half maximum) sizes of 2.2 arcsec, whereas similar scans across the bright stars α Car and β Ret yield FWHM sizes of 1.6 arcsec (two of the scans are shown in Fig. 1b); the reproducibility of these data implies uncertainties of ~ 0.1 arcsec.

The mid-infrared emission from SN1987A peaked on about day 120 and at that time most of the flux was photospheric, with a contribution from free–free emission². As expansion continued, the ejecta became optically thin at 10 μm and the flux, which became increasingly dominated by free–free emission, decreased, reaching a minimum around day 420 (Fig. 1c). Beyond this time, free–free and H^- emission will continue to decrease³, and cannot be responsible for the increase in 10 μm flux seen after about day 450. We attribute this increase to thermal emission from warm dust grains.

The 10 μm continuum flux increased markedly from 4 Jy on day 465 to 12 Jy on day 578. During the same period, the intensities of the emission lines have continued to decrease, and the broad emission hump at wavelengths shorter than 9.5 μm , attributed to SiO, is no longer readily discernible in the later

spectrum. Uncertainties in the contribution from SiO emission mean that the derived shape of the spectrum of the dust-emission component ranges from being featureless to being that corresponding to weak emission from silicate, such as might be expected from the wind of an oxygen-rich red supergiant which may have preceded the blue supergiant phase. Regardless of the exact shape of the emission, it has a characteristic colour temperature of ~ 400 K. A temperature of ≤ 400 K is also implied by the fact that the 1–4- μm flux has continued to fall (D. A. Allen, personal communication), because this indicates that the contribution from dust emission must be small at these wavelengths.

Whether the rise in 10- μm flux is due to the heating of pre-existing dust grains or to dust-grain condensation in the ejecta can be established by determining the location of the emitting dust. The 10- μm emission is extended with a FWHM of 1.5 ± 0.2 arcsec, corresponding to 0.38 pc at the distance of the Large Magellanic Cloud (which is considerably larger than the distance attained by even the highest-velocity material ejected by the supernova explosion). If the emitting grains were in the ejecta, they should condense in the dense material, which has velocities of the order of 2,000 km s^{-1} ; emission at the observed flux level, from material of the corresponding angular size at a temperature of 400 K, would be optically thick at 10 μm , and would be accompanied by substantial extinction at optical wavelengths. Also, a temperature of 400 K is considerably lower than that normally invoked as the condensation temperature of dust grains⁴. Thus, we conclude that the grains cannot have condensed in the supernova ejecta and that the increase in 10- μm flux arises from dust grains lying well beyond the ejecta, at a distance of ~ 1 light year from the supernova, which are heated by optical and ultraviolet light from the outburst. This phenomenon has been termed an infrared echo, because of the delay between the maximum emission at short wavelengths and the peak of infrared emission as seen by the observer^{5,6}. Because the integrated luminosity emitted by the supernova over the 50 days around the visible maximum (close to day 80) is about 50 times that estimated for the ultraviolet flash⁷, we will assume that it is the visible emission that dominates the heating process.

We now consider what the observed rate of increase of 10- μm flux and the spatial extent can reveal about the distribution of the emitting dust. At a time t measured from the observed explosion an observer may see infrared emission from within an ellipsoid of revolution with semi-latus-rectum ct and with the supernova and the observer at the foci. The light curve expected for an infrared echo has been discussed elsewhere^{6,8}, and it can be shown that, for a spherically symmetrical shell and for times t greater than the time of the visible maximum, the infrared luminosity can never increase faster than linearly with time and at about day 500 would be roughly constant or even diminishing⁶. In the present case no significant excess flux was detected before day 465, but since then the rate of increase with time has been much steeper than linear; the 10- μm flux has increased roughly as t^5 over the period day 465–578. This means that the simple model of a spherically symmetric distribution of emitting dust concentric with the supernova, and even asymmetric distributions such as ellipsoids⁸, can be ruled out, unless the visible light emission is itself highly anisotropic or the dust grains are unusually large (see below).

The linear extent of 1.5 ± 0.2 arcsec (FWHM) derived from the spatial scans is considerably smaller than the latus-rectum of the light ellipsoid on day 580, which would be ~ 3 arcsec 500 days after the visible maximum. Furthermore, the infrared profiles are centred closely on the visible supernova to within an accuracy of 0.25 arcsec. This effectively rules out other symmetric associations with the supernova, such as disks, toroids or bipolar configurations, and it appears that, for the flux to increase as steeply as observed, the light ellipsoid must not have reached the bulk of the material until $t \sim 400$ days after maximum. Most of the emitting material must lie behind the

‡ Present address: Royal Observatory Edinburgh, Edinburgh EH9 3HJ, UK.

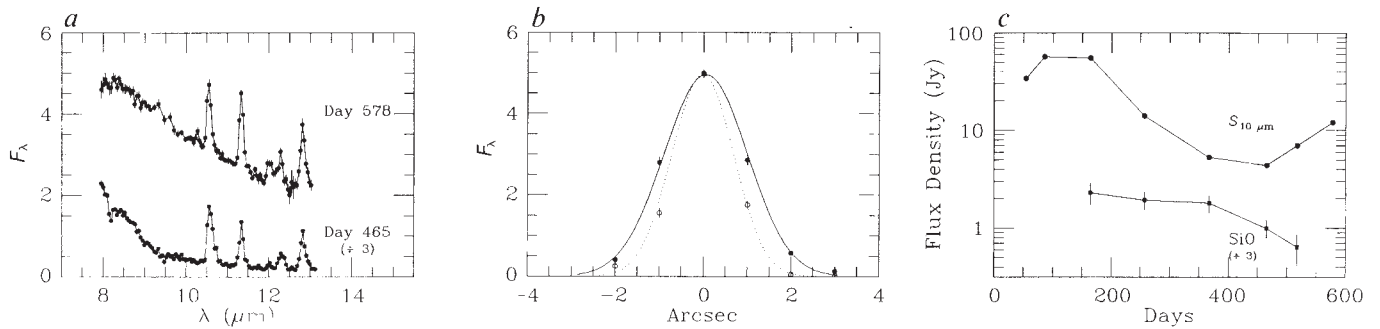


Fig. 1 The observations of SN1987A. *a*, 8–13- μm spectra on days 578 and 465. The latter spectrum has been divided by 3, so that the level of emission at 8–9.5 μm from SiO is roughly the same as that expected to be present in the day 578 spectrum. Emission lines of Co, Cl, Ni and Ne are prominent between 10 and 13 μm ; flux F_λ is in units of $10^{-17} \text{ W cm}^{-2} \mu\text{m}^{-1}$. *b*, Representative east-west scans across SN1987A and a reference star, β Ret, normalized to the peak on the supernova, at an effective wavelength of 9 μm . The solid and dashed lines are best-fit gaussian curves to the supernova (filled circles and solid line) and β Ret (open circles and dotted line) scans, and yield FWHM sizes of 2.3 and 1.7 arcsec respectively. The error bars plotted are statistical uncertainties only. *c*, The time development of the SiO emission (divided by 3) and the $10\text{-}\mu\text{m}$ continuum flux.

supernova and at a distance of ~ 200 light days or 0.2 parsec (Fig. 2). As this material is so close to the supernova it is probably part of the outflow from the precursor star. If we assume that the supernova light emission is intrinsically isotropic, then its outflow has become somehow severely distorted, perhaps by the motion of the precursor through the ambient interstellar medium or alternatively by interaction with the outflow from a luminous H II region complex lying behind the supernova.

One can calculate the temperature attained by dust grains heated by the light emitted during the optical maximum of the supernova. At a distance of 0.3 pc from the supernova, which had, at maximum, an effective temperature of 5,500 K and a luminosity of $2.5 \times 10^8 L_\odot$ (ref. 9), dust grains with radius $a = 0.1 \mu\text{m}$ and emission efficiency proportional to $1/\lambda^2$ in the infrared (where λ is the wavelength) would reach a temperature of about 400 K (ref. 10), in agreement with that observed. The mass of emitting grains then follows from

$$S_{10\mu\text{m}} = B(10 \mu\text{m}, 400 \text{ K}) \kappa M_d / d^2$$

where d is the distance to the supernova, κ is the mass absorption coefficient of the dust of total mass M_d (3,000 and $360 \text{ cm}^2 \text{ g}^{-1}$ for silicate and graphite grains respectively at $10 \mu\text{m}$ (ref. 11)), S is the emitted flux and B is the Planck function. Assuming that $d = 52 \text{ kpc}$ and that the excess emission from dust on day 580 is 10 Jy, then $M_d = 4 \times 10^{-5} M_\odot$ for silicates and is larger by a factor of eight for graphite grains. In the discussion below, we assume that the grains are silicates; if instead carbon grains are assumed, the quantities derived must be increased by a factor eight.

The volume in which this mass of emitting dust is contained can be estimated from the observed spatial extent, r , and a depth, l , derived from the duration of the broad maximum in the optical emission from the supernova (~ 50 days). This volume is $\pi r^2 l = 1.4 \times 10^{53} \text{ cm}^3$. The gas-to-dust ratio in the Magellanic Clouds is likely to exceed the value of ~ 200 for our Galaxy and so the number density of dust grains implies a hydrogen density $\geq 50 \text{ cm}^{-3}$.

The luminosity of the infrared excess on day 578 is $\sim 2 \times 10^5 L_\odot$, and is due to the interception of $\sim \pi/2$ sr of the supernova output, thus representing a visible optical depth of $\sim 1\%$ compared with the early optical luminosity. This rules out the possibility that the initial ultraviolet flash caused a large fraction of the infrared emission, as the lower integrated luminosity of the flash would lead to very high optical depths and ultimately to an energy-balance problem. Similarly, the resolved size and the temperature of 400 K imply an optical depth at $10 \mu\text{m}$ of 5×10^{-4} , which is consistent with the ratio of extinction at visible and infrared wavelengths.

Emission from a putative spherical fossil shell of radius 0.2 pc would be close to the end of its plateau phase by around day 580⁷. Such emission at $10 \mu\text{m}$ must at all times have been less than 4 Jy, which is approximately the minimum flux observed (the infrared excess emission above the photosphere seen before day 400 is due to free-free processes^{2,3}). This places constraints on the density of dust within the light ellipsoid and suggests either that the precursor red giant did not form dust or, if the dust/gas ratio is 'normal', that the gas density in the inner zones of the shell is substantially less than the 50 cm^{-3} value derived from the emission on day 580. Thus

$$n_H = \frac{\dot{M}}{4\pi v^3 t^2 m_H} \leq 10 \text{ cm}^{-3}$$

where t is the time since the end of the red supergiant phase, v is the outflow velocity, $\dot{M}$ is the mass loss rate, n_H is the density of hydrogen and m_H is the mass of a hydrogen atom. If $v \approx 20 \text{ km s}^{-1}$ and $t \approx 10^4 \text{ yr}$, then the mass loss rate during the red supergiant phase is $\dot{M} \leq 2 \times 10^{-6} M_\odot \text{ yr}^{-1}$.

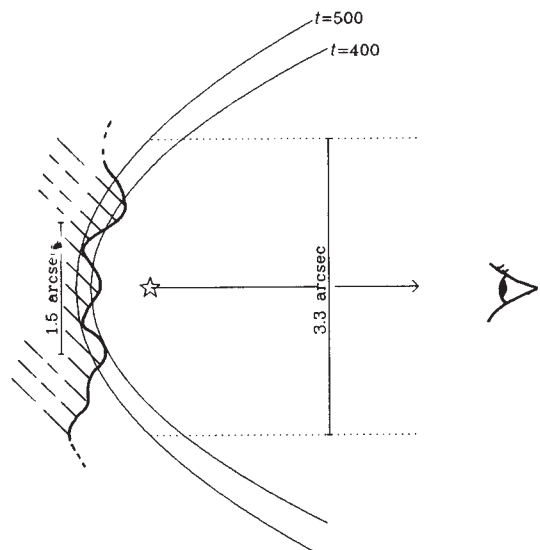


Fig. 2 A sketch showing the probable positions of the supernova and the emitting dust relative to the observer. The curves represent the extent of the light ellipsoids 400 and 500 days after the optical maximum. The bars represent the FWHM extent of the $9\text{-}\mu\text{m}$ emission derived from the spatial scans, and the hatched area denotes the location of the emitting dust.

As a possible alternative model which retains spherical symmetry, we point out that if the grains are large ($a \geq 100 \mu\text{m}$), they could support internal temperature gradients such that only the side of the grain facing the star is heated sufficiently to radiate at $10 \mu\text{m}$. In such a model, the close alignment of the centroid of the dust emission to the supernova would be explained naturally. The individual grains would be optically thick at $10 \mu\text{m}$, and would therefore give rise to only a very weak silicate feature. Apart from the need for virtually complete destruction of the smaller grains, however, the main argument against this picture is that a large mass of dust ($\sim 0.04 M_{\odot}$) is required to produce the observed $10\text{-}\mu\text{m}$ flux, for radii $a = 100 \mu\text{m}$ and temperature $T = 400 \text{ K}$.

If the emission continues to brighten, it should become possible to define the spectral shape of the dust emission more accurately. As the light ellipsoid penetrates further into the material, the emission should become more spatially extended. Continued monitoring of the infrared emission from SN1987A should reveal more details of the material lying on the far side of the supernova.

Received 2 November 1988; accepted 4 January 1989.

1. Smith, C. H., James, S. D., Aitken, D. K., Orton, G. & Roche P. F. *JAU Circ.* No. 4645 (1988).
2. Aitken, D. K., et al. *Mon. Not. R. astr. Soc.* **231**, 7P-14P (1988).
3. Aitken, D. K. et al. *Mon. Not. R. astr. Soc.* **235**, 19P-31P (1988).
4. Draine, B. T. in *Physical Processes in Red Giants* (eds Iben, I. & Renzini, A.) 317-333 (Reidel, Dordrecht, 1981).
5. Graham, J. R. et al. *Nature* **304**, 709-710 (1983).
6. Dwek, E. *Astrophys. J.* **274**, 175-183 (1983).
7. Kirscher, R. P. in *ESO Conf. Workshop Proc.* **26**, 121-131 (1987).
8. Emmering R. T. & Chevalier, R. A. *Astr. J.* **95**, 152-157 (1988).
9. Catchpole, R. M. et al. *Mon. Not. R. astr. Soc.* **229**, 15P-25P (1988).
10. Dwek, E., Sellgren, K., Soifer, B. T. & Werner, M. W. *Astrophys. J.* **238**, 140-147 (1980).
11. Draine, B. T. *Astrophys. J. Suppl.* **57**, 587-594 (1985).

Evidence for a decline in the atmospheric accumulation rate of CHClF_2 (CFC-22)

Curtis P. Rinsland*, Douglas W. Johnson†, Aaron Goldman‡ & Joel S. Levine*

* Atmospheric Sciences Division, NASA Langley Research Center, Hampton, Virginia 23665-5225, USA

† Battelle, Columbus Division, 505 King Avenue, Columbus, Ohio 43201-2693, USA

‡ Department of Physics, University of Denver, Denver, Colorado 80208-0202, USA

Concern that emissions of fully halogenated hydrocarbons may deplete stratospheric ozone through chlorine-catalysed destruction¹⁻³ has prompted a search for safer chemicals. CHClF_2 (CFC-22) is a promising substitute as it contains only a single chlorine atom and is partially removed in the troposphere by reactions with hydroxyl radicals (OH)⁴. Recognizing these facts, CFC-22 has been excluded from the controls under the Montreal Protocol²⁹. The possibility of rapidly increasing use of CFC-22 suggests a need for careful atmospheric monitoring, especially as the atmospheric lifetime of CFC-22 is long but uncertain (12-28 yr (refs 3, 5)) and absorption by the strong CFC-22 infrared bands could contribute to greenhouse warming^{6,7}. Here we report atmospheric CFC-22 measurements derived from ground-based solar spectra recorded between December 1980 and May 1988 which show that the CFC-22 total column increased at an average annual exponential rate of $7.8\% \pm 1.0\%$ (2σ). Compared with other atmospheric data, these measurements indicate that CFC-22 is increasing at a more rapid rate than either CFC-11 or CFC-12, the two most abundant chlorofluorocarbons, but that the rate of CFC-22 increase is likely to have declined over the past few years.

CHClF_2 (chlorodifluoromethane, or CFC-22) is a man-made

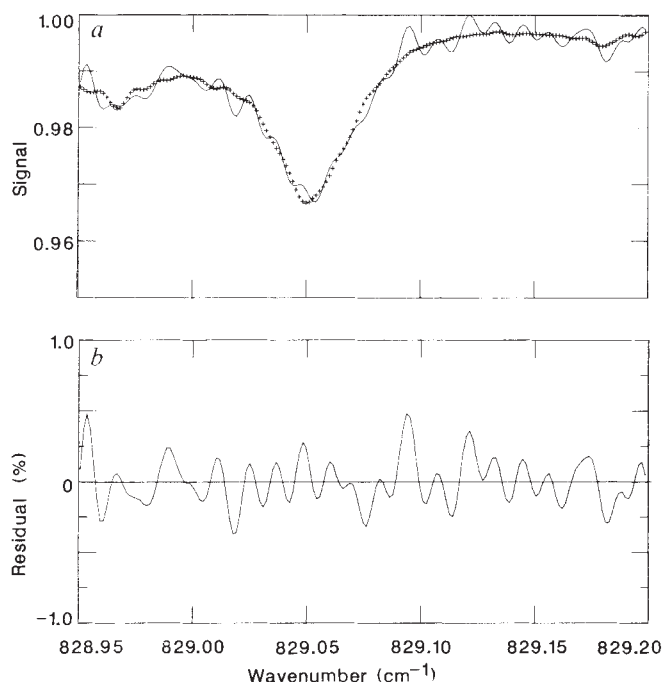


Fig. 1 *a*, Comparison between a Kitt Peak solar spectrum recorded at 0.0050 cm^{-1} resolution and a mean solar astronomical zenith angle of 80.47° on 6 October 1982 (solid line), and a least-squares best fit to the data (+). *b*, Residuals (observed minus calculated values). The signal-to-r.m.s. noise ratio of this spectrum (~ 500) is typical of the Kitt Peak data for the CFC-22 Q-branch region. In addition to the CFC-22 parameters, the line parameters on the 1986 Air Force Geophysics Laboratory compilation²⁷ were included. Based on a set of reference VMR profiles, primarily those of Smith²⁸, only minor contamination is predicted from weak CO_2 lines at 828.9691 and 829.0718 cm^{-1} and from several overlapping weak lines of C_2H_6 in the interval $828.91\text{--}828.97 \text{ cm}^{-1}$, and this can be identified in the fitted region ($828.95\text{--}829.20 \text{ cm}^{-1}$). A weak, high-frequency channel spectrum is present in the Kitt Peak spectrum (0.1% amplitude, 0.118 cm^{-1} period). A least-squares analysis that includes fitting of this channel spectrum yields a change in the retrieved CFC-22 total column of only 0.1% .

halocarbon with a current tropospheric abundance of about one fourth that of CFC-12 (CCl_2F_2) and one third that of CFC-11 (CCl_3F). CFC-22 is primarily used in refrigeration, but a significant portion, about 35% of total production, is used in fluoropolymer production, which results in little emission of CFC-22 into the atmosphere⁸. Increased use of fluoropolymers is believed to account for much of the recent increase in CFC-22 production⁸. The quantitative history of CFC-22 release into the atmosphere is poorly known. Surface-level mixing ratios (see, for example, refs 9-11) and vertical profiles in the troposphere and stratosphere¹² have been found to be significantly higher than values calculated from reported industrial release rates^{3,9-12}, which indicates that those rates are either underestimated or incomplete, that there exist additional sources of atmospheric CFC-22, that the measured mixing ratios are too high, or that a combination of these effects is occurring. The measured surface-level concentrations exceed the calculated values even if no destruction is assumed in both the troposphere and stratosphere^{9,10}, so that at present it is impossible to estimate reliably the CFC-22 atmospheric lifetime from emission and measurement data.

To derive the long-term trend in the total column of atmospheric CFC-22, we have analysed absorption by the 829.05-cm^{-1} $2\nu_2$ Q-branch of CFC-22 in high-resolution solar spectra recorded on nine days between December 1980 and May 1988. The spectra were obtained with the US National Solar Observatory (NSO) Fourier transform spectrometer located on Kitt Peak

(latitude 31.9° N, longitude 111.6° W, elevation 2.095 km). The spectral resolution, defined as $1/2\sigma_{\max}$ where $\sigma_{\max}$ is the maximum path difference, ranges from 0.0025 to 0.0053 cm^{-1} .

Total CFC-22 vertical columns above Kitt Peak have been determined using the technique of nonlinear least-squares spectral fitting¹³⁻¹⁵. A total of four parameters were adjusted simultaneously in fitting each spectrum: two parameters model the level and slope of the 100% transmission level in the analysis region; a single parameter is used to determine the frequency offset (cm^{-1}) between the measured and calculated spectra; and the fourth parameter is the multiplicative scale factor that converts an assumed CFC-22 relative volume-mixing ratio (VMR) profile to an absolute profile. For the relative profile we adopted values measured over the 10–30-km altitude range and, in the troposphere, a constant VMR equal to the measured value of 55 parts per 10^{12} by volume (p.p.t.v.) at 10 km (ref. 16). Correlative pressure-temperature profiles from radiosonde soundings at the Tucson airport on the day of the spectral measurements or US National Meteorological Center profiles from global satellite and radiosonde soundings (K. Johnson and M. Gelman, personal communication) were assumed in the ray-tracing calculations.

Although the 829.05- cm^{-1} Q-branch was first used for the spectroscopic detection of CFC-22 in the atmosphere over seven years ago¹⁷, no line-by-line parameters have been reported for this feature. As in an earlier study¹⁶, we had to use empirical parameters deduced by fitting reference laboratory spectra. A file containing 626 lines evenly spaced in wavenumber between 828.525 and 829.895 cm^{-1} was created by matching the shape and intensity of the Q-branch absorption in a 0.02- cm^{-1} resolution laboratory spectrum recorded at 23 °C (ref. 18). The integrated Q-branch intensity of $3.33 \times 10^{-19} \text{ cm}^{-1}$ per molecule per cm^2 at 296 K between 828.955 and 829.141 cm^{-1} agrees well with the value of $3.60 \times 10^{-19} \text{ cm}^{-1}$ per molecule per cm^2 derived for the same interval from other laboratory measurements¹⁶. The earlier qualitative estimate¹⁷ of $\sim 1 \times 10^{-19} \text{ cm}^{-1}$ per molecule per cm^2 provides a lower limit, as it assumes linear absorption in deriving the total Q-branch intensity from the laboratory spectra. We adopted an empirically deduced lower-state energy of 230 cm^{-1} for all lines¹⁶. A mean air-broadened halfwidth was derived by requiring that the calculated Q-branch shape match the shape measured in the solar spectra and that all solar spectra recorded on the same day yield CFC-22 total columns that agree within their estimated precisions. The temperature dependence of the CFC-22 line intensities was calculated assuming, for the rotational partition function, the standard $T^{1.5}$ temperature factor, and for the vibrational partition functions, those computed using the harmonic-oscillator approximation¹⁹ and the fundamental frequencies reported by Plyler and Benedict²⁰.

The spectral retrievals (an example of which is shown in Fig. 1) result in the daily-averaged CFC-22 total column amounts and fitted trend shown in Fig. 2. Error bars (5–10%) in Fig. 2 are estimated 2σ precision limits based on the spectral signal-to-noise ratio, the agreement among the various determinations, and the number of observations on each day. Absolute error limits are much larger ($\sim 25\%$), primarily because the assumed CFC-22 spectral parameters are based on limited laboratory data. The error sources and estimated uncertainties in the absolute CFC-22 total column due to each source are instrument effects (noise, for example) ($\pm 2\%$), and uncertainties in the CFC-22 spectroscopic parameters at room temperature ($\pm 15\%$), the temperature corrections to these parameters ($\pm 20\%$), the simulation of interfering lines ($\pm 2\%$), and the assumed vertical CFC-22 distribution ($\pm 4\%$).

With the assumed vertical VMR distribution, the corresponding mean tropospheric CFC-22 VMRs are 38 ± 10 p.p.t.v. on 21 December 1980 and 68 ± 17 p.p.t.v. on 23 May 1988, which are 30–40% lower than surface-level VMRs measured in the Northern Hemisphere^{9-11,21,22}. Similarly, all other infrared measure-

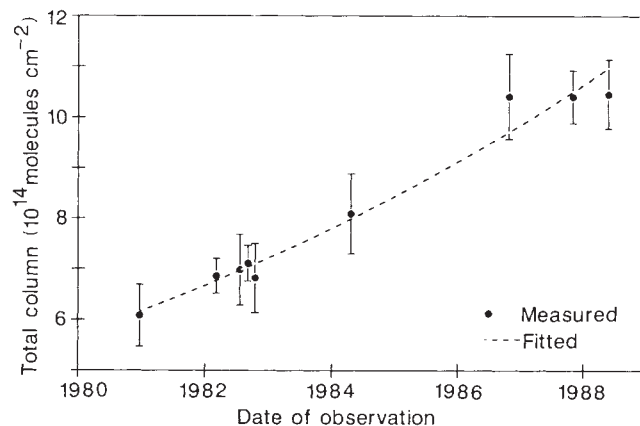


Fig. 2 Daily-averaged CFC-22 total column amounts above Kitt Peak plotted against observation date. Error bars show 2σ precisions estimated for each day. The dashed curve is the trend derived from the data, and corresponds to an increase by a factor of 1.78 in the CFC-22 total column (in molecules cm^{-2}), from 6.16×10^{14} on 21 December 1980 to 1.098×10^{15} on 23 May 1988.

ments^{16,23}, including values^{17,24} corrected to account for the low integrated Q-branch intensity of ref. 17, are lower than corresponding air-sampling measurements. Although the infrared VMRs are closer to values calculated from estimated industrial release rates^{3,9-12}, the systematic difference between the infrared and air-sampling data needs to be resolved before historical CFC-22 emissions and atmospheric lifetime estimates can be reliably tested by observations. Both the infrared spectroscopic parameters and the air-sampling absolute calibration (which is uncertain for at least some data²⁵) require further verification.

Our results indicate a $7.8\% (\pm 10\% 2\sigma) \text{ yr}^{-1}$ average exponential rate of increase in the CFC-22 total column above Kitt Peak from December 1980 to May 1988, derived by fitting the daily-averaged measurements to the formula $(1/C) dC/dt = \beta$, where C is the measured total column and β is the rate of increase. There are too few measurements to detect seasonal variations or changes in the rate of increase over this time period. Because of the long lifetime of CFC-22, our measured increase rate should be representative of the global trend; it is substantially higher than corresponding values for CFC-11 and CFC-12 ($\sim 5\% \text{ yr}^{-1}$)⁸.

Electron-capture-gas-chromatographic analysis of surface-level air samples from the US Pacific Northwest (45° N latitude) and the South Pole yielded average exponential CFC-22 increase rates of 10–13% yr^{-1} before 1983^{10,21,22}, consistent with a summary of reported surface measurements from 1976 to 1982¹¹. Our slower rate, derived from data up to May 1988, suggests a significant recent decrease in the per cent rate of atmospheric CFC-22 accumulation. This conclusion is supported by measured surface-level CFC-22 increases of 8% yr^{-1} at Cape Grim, Tasmania (41° S), between August 1984 and December 1985²⁶, and 7.1% yr^{-1} for 1986²⁵, and also by the slower growth of industrial production estimated for the present decade ($\sim 8\%$ for 1981–1984⁸, about half that for the 1970s^{8,10}). It is less consistent, however, with the higher rate of increase of 10% determined for CFC-22 by comparing balloon-borne profiles from 1982–83 with those for 1987¹². The declining rate of atmospheric accumulation noted here is likely to reverse in the future, as CFC-22 is substituted for CFC-11 and CFC-12 in response to the Montreal Protocol. However, the discrepancy noted here between the infrared and *in situ* CFC-22 measurement needs to be resolved before the absolute values of these data can be reliably considered in budget studies.

We thank Claude Plymate, Jeremy Wagner, Ron Hubbard and Michael Brown for their assistance in recording the solar spectra, Gregg Ladd for computer processing of the data, Jack

Fishman for helpful suggestions, and M. A. K. Khalil and P. Fabian for sending us recently published CFC-22 measurements. The Kitt Peak observations were partially supported by the US Department of Energy CO₂ program, NASA, the Chemical Manufacturers Association and NSF.

Received 28 September; accepted 12 December 1988.

1. Molina, M. J. & Rowland, F. S. *Nature* **249**, 810–812 (1974).
2. Rowland, F. S. & Molina, M. J. *Rev. Geophys. Space Phys.* **13**, 1–35 (1975).
3. Wuebbles, D. J. *J. geophys. Res.* **88**, 1433–1443 (1983).
4. Natn. Acad. Sci. *Stratospheric Ozone Depletion by Halocarbons: Chemistry and Transport* (Natn. Acad. Sci., Washington, DC, 1979).
5. World Met. Org. Rep. No. 11, Geneva (1981).
6. Ramanathan, V., Cicerone, R. J., Singh, H. B. & Kiehl, J. T. *J. geophys. Res.* **90**, 5547–5566 (1985).
7. Varanasi, P. & Chudamani, S. J. *geophys. Res.* **93**, 1666–1668 (1988).
8. WMO/NASA Atmospheric Ozone 1985: Assessment of Our Understanding of the Processes Controlling its Present Distribution and Change World Met. Org. Global Research and Monitoring Proj. Rep. No. 16, Ch. 3 (1986).
9. Rasmussen, R. A., Khalil, M. A. K., Penkett, S. A. & Prosser, N. J. D. *Geophys. Res. Lett.* **7**, 809–812 (1980).
10. Khalil, M. A. K. & Rasmussen, R. A. *Nature* **292**, 823–824 (1981).
11. Fabian, P. in *The Handbook of Environmental Chemistry*, Vol. 4/Part A (ed. Hutzinger, O.) 24–51 (Springer, Berlin & Heidelberg, 1986).
12. Fabian, P. et al. in *Proc. Quadrennial Ozone Symp.* Göttingen, West Germany (in the press).
13. Rinsland, C. P. et al. *J. geophys. Res.* **87**, 11119–11125 (1982).
14. Rinsland, C. P., Boughner, R. E., Larsen, J. C., Stokes, G. M. & Brault, J. W. *J. geophys. Res.* **89**, 9613–9622 (1984).
15. Rinsland, C. P., Levine, J. S. & Miles, T. *Nature* **318**, 245–249 (1985).
16. Zander, R., Rinsland, C. P., Farmer, C. B. & Norton, R. H. *J. geophys. Res.* **92**, 9836–9850 (1987).
17. Goldman, A. et al. *Geophys. Res. Lett.* **8**, 1012–1014 (1981).
18. Murcray, D. G., Murcray, F. J., Goldman, A., Bonomo, F. S. & Blatherwick, R. D. *High-Resolution Infrared Laboratory Spectra* (Univ. Denver, 1984).
19. Herzberg, G. *Molecular Spectra and Molecular Structure. II. Infrared and Raman Spectra of Polyatomic Molecules* (Van Nostrand Reinhold, New York, 1945).
20. Plyler, E. K. & Benedict, W. S. *J. Res. Natn. Bur. Stand.* **47**, 202–220 (1951).
21. Rasmussen, R. A. & Khalil, M. A. K. *Antarct. J. U.S.* **17**, 203–205 (1982).
22. Rasmussen, R. A. & Khalil, M. A. K. *Antarct. J. U.S.* **18**, 250–252 (1983).
23. Rinsland, C. P. et al. *J. geophys. Res.* **93**, 12607–12626 (1988).
24. Zander, R., Stokes, G. M. & Brault, J. W. *Geophys. Res. Lett.* **10**, 521–524 (1983).
25. Watson, R. T. et al. *NASA Ref. Publ.* 1208, Table C-8.1 (1988).
26. Khalil, M. A. K. & Rasmussen, R. A. in *Baseline Atmospheric Program (Australia) 1985* (eds Forgan, B. W. & Fraser, P. J.) 26–29 (Dept. Sci./Bur. Met., Australia, 1987).
27. Rothman, L. S. et al. *Appl. Opt.* **26**, 4058–4097 (1987).
28. Smith, M. A. H. *NASA tech. Mem.* 83289 (1982).
29. Johnston, K. *Nature* **329**, 189 (1987).

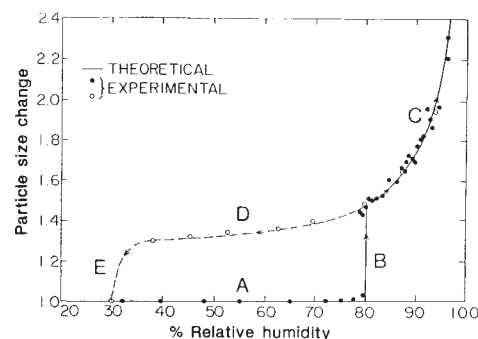


Fig. 1 Dependence of particle size on relative humidity for a particle of $(\text{NH}_4)_2\text{SO}_4$ (ref. 17). Closed and open circles represent experimental data, and signify increasing and decreasing relative-humidity conditions, respectively. The solid line is the result of theoretical calculations. Reprinted with permission of the author.

energy, mass and volume¹⁰. A system may remain at this level of energy for arbitrarily long times, even though more stable thermodynamic states exist.

For example, supercooled fog droplets are in metastable equilibrium with respect to ice but are stable with respect to H₂O vapour. Atmospheric aerosols can also exist in metastable equilibrium with respect to their H₂O content. Aerosol particles that exist as droplets at relative humidity (RH) values below the deliquescence humidity (DH) of their solute are in metastable equilibrium with respect to the solid phase.

Aerosol particles existing in metastable equilibrium as droplets, instead of the more thermodynamically stable solid phase, are larger than their dry counterparts¹. The existence of larger droplets gives rise to atmospheric phenomena such as visibility degradation^{2,3}. Increased particle size enhances gas- and liquid-phase mass-transfer-limited chemical reactions, which depend on the surface area of the aerosol particles⁴. The existence of a dispersed aqueous phase can also have a significant effect on the oxidation of SO₂ and on the equilibrium vapour pressures of SO₂, NH₃, N₂O₅, HNO₂ and H₂O₂ (refs 5–9).

Atmospheric aerosol particles with diameters <2 μm consist primarily of inorganic compounds such as $(\text{NH}_4)_2\text{SO}_4$, NaCl and NH_4NO_3 ^{11,12}. Such particles will exhibit a number of properties that are dependent on RH, such as particle size, phase, liquid H₂O content and hysteretic phenomena involving metastable states^{13–18}.

For example, consider a solid deliquescent particle such as $(\text{NH}_4)_2\text{SO}_4$. The particle is solid at sufficiently low RH (Fig. 1, curve A), but when the RH increases above the particle's DH, the particle changes phase to a droplet of aqueous solution as H₂O vapour condenses rapidly onto its surface. Condensation of H₂O vapour continues until the particle is saturated with respect to solute concentration (Fig. 1, curve B). Further increase in RH then results in dilution and undersaturation of the solute and an increase in particle size (Fig. 1, curve C). Reducing the RH below the particle's DH leaves it in a metastable state, causing hysteresis; the particle remains a droplet but becomes supersaturated with respect to solute concentration (Fig. 1, curve D). Further decrease in RH increases the solute concentration, eventually to a degree that is sufficient to initiate homogenous or heterogeneous nucleation of crystals of solute within the droplet and evaporation of liquid H₂O (efflorescence; Fig. 1, curve E).

The methodology of detecting metastable ambient aerosols has been described elsewhere^{19–21}. Ambient aerosol is drawn continuously through a system that rapidly heats, cools, and then detects the continuous-flow sample with an integrating nephelometer. The first heat exchanger scans the sample temperature from ambient temperature to 350 °C within 4 minutes, and the second heat exchanger cools the sample back to ambient

Ubiquitous nature of ambient metastable aerosol

M. J. Rood*, M. A. Shaw*, T. V. Larson†
& D. S. Covert†

* University of Illinois at Urbana-Champaign,
Department of Civil Engineering, Urbana,
Illinois 61801, USA

† University of Washington, Department of Civil Engineering,
Seattle, Washington 98195, USA

The potential for atmospheric aerosol particles to exist as metastable aqueous droplets instead of in their thermodynamically favoured crystalline solid phase has many implications for atmospheric physics and chemistry. Aqueous droplets are larger than their dry counterparts¹ and generally scatter more light^{2,3}. Also, the existence of an aqueous phase influences the oxidation, equilibrium vapour pressure and mass-transfer kinetics of trace gases^{4–9}. Here we report observations at three urban and rural sites, showing that metastable droplets existed more than 50% of the time when the ambient relative humidity was between ~45 and 75%. At one site, more than 10% of the aerosol's light-scattering coefficient was attributed to metastable liquid H₂O when the relative humidity was greater than 55%. These results indicate that metastable droplets are ubiquitous in nature, although the full ramifications of these observations are still unclear.

Metastable equilibrium occurs at a local minimum of the Gibbs free energy, where the second virtual variation in entropy with respect to neighbouring states is zero at constant internal

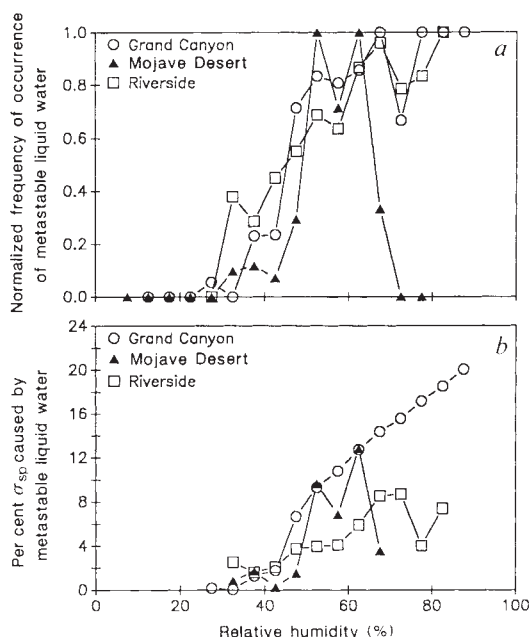


Fig. 2 Dependence of the frequency of occurrence of metastable liquid H_2O (a) and the normalized light-scattering extinction coefficient associated with metastable liquid H_2O (b) on ambient relative humidity.

conditions. The light-scattering coefficient of the aerosol particles (σ_{sp}) is then measured at ambient temperature and RH. The aerosol is considered to contain metastable droplets if there is an abrupt decrease in σ_{sp} of at least 2% within the first 30 °C of temperature scan; this is chosen because it is discernable from the noise present during temperature scans. Such a sharp decrease in σ_{sp} occurs for metastable droplets because the particles effloresce (Fig. 1, curve E) as they experience low-RH conditions at increased heater temperatures. On cooling, the particles do not regain liquid H_2O associated with the metastable solute because the final RH of the sample is less than the DH of the solute. Particles undersaturated with respect to solute concentration (Fig. 1, curve C) at ambient RH conditions do not exhibit a change in σ_{sp} on heating the sample to 30 °C above ambient conditions because the liquid H_2O that evaporates from the droplets in the heater recondenses on cooling to ambient conditions; dry particles (Fig. 1, curve A) do not exhibit such a change because thermal decomposition of the solute does not occur until heater temperatures are at least 30 °C above ambient conditions.

The general nature of ambient metastable droplets was observed experimentally at three locations: Riverside, California; Mojave Desert, California; and Grand Canyon, Arizona (Table 1). Riverside is an urban site which is influenced by maritime and continental air masses. Mojave Desert and Grand Canyon are rural sites dominated by continental air masses. The

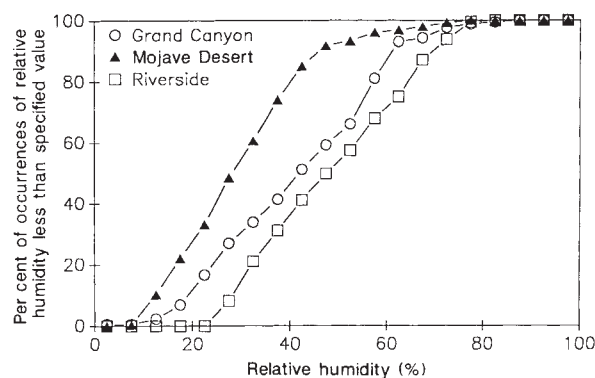


Fig. 3 Cumulative distribution of ambient relative-humidity conditions during the specified experiments.

magnitude of ambient σ_{sp} values are in general agreement with experimental values reported previously from the Grand Canyon^{22,23} and Riverside²⁴ areas. All sampling occurred between the summer months of July and September. A seasonal dependence of the results is expected only if such variations would cause freezing of the aerosol droplets.

The frequency of occurrence of ambient metastable liquid H_2O depends on RH and the particles' deliquescence and efflorescence humidities (Fig. 2a). Rapid increases and decreases in this frequency of occurrence were observed at ~30% and 80% RH respectively (Fig. 2a). The existence of metastable droplets did not appear to depend on whether ambient RH was increasing or decreasing. These RH values are in approximate agreement with the efflorescence and deliquescence humidities, respectively, of compounds such as $(\text{NH}_4)_2\text{SO}_4$ and NaCl (ref. 17). Metastable droplets existed more than 50% of the time for ambient RH values between ~45 and 75%. Such a response is significant because ambient RH values for all three sites were in this range 38% of the time (Fig. 3).

The amount of light scattered by metastable liquid H_2O is important because σ_{sp} is proportional to the mass concentration of atmospheric aerosol particles with diameters $<2 \mu\text{m}$ (ref. 25). The amount of light scattered by metastable liquid H_2O increases with RH for RH ≈ 30 –80% (Fig. 2b). At least 5% of the scattering was attributed to metastable liquid H_2O for RH values between ~45 and 75%. Scattering attributed to metastable liquid H_2O at Grand Canyon accounted for more than 10% of the total σ_{sp} when the RH was $>55\%$.

The formation and existence of ambient metastable aerosol depends on the chemical and physical history of the aerosol. A hygroscopic, acidic aerosol particle can exist as a droplet at moderate RH values. Such an acidic droplet could react with a gaseous base to form an aqueous droplet with a DH greater than ambient RH conditions (for example, the reaction $\text{H}_2\text{SO}_4 + 2\text{NH}_3 \rightarrow (\text{NH}_4)_2\text{SO}_4$ occurring at 50% RH). Spatial and/or temporal changes in RH during transport of an aerosol could result first in a deliquescent particle experiencing RH values above

Table 1 Summary of experimental results

Location	Grand Canyon	Mojave Desert	Riverside
Total number of temperature scans	205	266	217
Per cent of temperature scans used in the data analysis	85	96	96
Per cent of tests indicating metastable droplets	38	10	57
Mean σ_{sp} value (10^{-5} m^{-1})*	1.05	2.61	25.1
Standard deviation (10^{-5} m^{-1})	5.97	1.58	11.6
Dates of sampling	31/7/84–22/8/84	29/8/84–21/9/84	13/9/84–18/9/84
Elevation of sampling site (m)	1,820	790	360

* For particles with diameters $<2 \mu\text{m}$.

the DH of its solute, thus forming a droplet, which might then encounter RH values less than the DH of the solute, resulting in the formation of a metastable droplet.

Our results indicate that ambient metastable aerosol droplets are common in the atmosphere. At RH values between ~30% and 80%, metastable liquid H₂O contributes significantly to ambient aerosol liquid H₂O content and mass concentration of atmospheric aerosols. The full consequences for atmospheric thermodynamics and visibility degradation remain to be determined.

The authors are grateful for cooperation and field support from Dr J. N. Pitts, Jr and Dr A. P. Waggoner. This research was supported by the NSF.

Received 12 September; accepted 12 December 1988.

- Orr, C., Hurd, F. K. & Corbett, W. J. *J. Colloid Sci.* **13**, 472-482 (1958).
- Toon, O. B. & Pollack, J. B. *J. appl. Met.* **15**, 225-246 (1976).
- Sloane, C. S. *Atmos. Envir.* **18**, 871-878 (1984).
- Larson, T. V. & Taylor, G. S. *Atmos. Envir.* **17**, 2489-2495 (1983).
- Freiberg, J. *Envir. Sci. Technol.* **8**, 731-734 (1974).
- Cadle, R. D. & Robbins, R. C. *Discuss. Faraday Soc.* **30**, 155-166 (1961).
- Brock, J. R. & Durham, J. L. in *SO₂, NO and NO₂ Oxidation Mechanisms: Atmospheric Considerations* (ed. Calvert, J. G.) 209-249 (Ann Arbor Science, Boston, 1984).
- Hoffman, M. R. & Jacob, D. J. in *SO₂, NO and NO₂ Oxidation Mechanisms: Atmospheric Considerations* (ed. Calvert, J. G.) 101-172 (Ann Arbor Science, Boston, 1984).
- Penkett, S. A., Jones, B. M. R., Brice, K. A. & Eggleton, A. E. *J. Atmos. Envir.* **13**, 123-137 (1979).
- Pruppacher, H. R. & Klett, J. D. *Microphysics of Clouds and Precipitation* 73 (Reidel, London, 1980).
- Brosset, C. *Atmos. Envir.* **12**, 25-38 (1978).
- Ono, A. *Atmos. Envir.* **12**, 753-757 (1978).
- Dessens, H. Q. *J. R. met. Soc.* **75**, 23-26 (1949).
- Junge, C. *Air Chemistry and Radioactivity* 133 (Academic Press, New York, 1963).
- Winkler, P. *Aerosol Sci.* **4**, 373-387 (1973).
- Hänel, G. *Adv. Geophys.* **3**, 73-188 (1976).
- Tang, I. N. in *Generation of Aerosols and Facilities for Exposure Experiments* (ed. Willeke, K.) 153-167 (Ann Arbor Science, Ann Arbor, 1980).
- Spann, J. F. & Richardson, C. B. *Atmos. Envir.* **19**, 819-825 (1985).
- Larson, T. V., Ahlquist, N. C., Weiss, R. E., Covert, D. S. & Waggoner, A. P. *Atmos. Envir.* **16**, 1587-1590 (1982).
- Rood, M. J., Covert, D. S. & Larson, T. V. *Tellus* **39B**, 383-397 (1987).
- Rood, M. J., Covert, D. S. & Larson, T. V. *Aerosol Sci. Technol.* **7**, 57-65 (1987).
- Malm, W. C. & Molenaar, J. V. *J. Air Pollut. Control Ass.* **34** (9) 899-904 (1984).
- Ferek, R. J., Farber, R. J. & Bhardwaja, P. S. in *Trans. Visibility Conf., Research and Policy Aspects* (ed. Bhardwaja, P. S.) 57-75 (1987).
- Larson, S. M. & Cass, G. R. *Envir. Sci. Technol.* (in the press).
- Waggoner, A. P. & Weiss, R. E. *Atmos. Envir.* **14**, 623-626 (1980).

Proton environments and hydrogen-bonding in hydrous silicate glasses from proton NMR

S. C. Kohn*, R. Dupree† & M. E. Smith†

* Department of Geology, University of Manchester, Oxford Road, Manchester M13 9PL, UK

† Department of Physics, University of Warwick, Coventry CV4 7AL, UK

The dissolution of water in magmas has profound effects on their physical and chemical properties, and an understanding of this process continues to be of major importance in igneous petrology and geochemistry. Systematic studies of the physical properties of hydrous melts, and infrared^{1,2}, Raman^{2,3} and ²⁹Si NMR⁴ spectroscopic studies of glasses quenched from melts, have provided valuable information, but many questions remain unanswered. Previous proton NMR studies of glasses have generally concentrated on lineshape, to distinguish between the proton environments in Si-OH groups and in molecular H₂O (refs 5, 6). Here we present proton NMR results in glasses quenched from hydrous melts, which allow us to resolve different proton sites on the basis of chemical-shift differences. Four separate proton resonances are resolved in a silica glass containing 8.7 wt% H₂O, and a strongly hydrogen-bonded Si-OH group is identified in hydrous alkali and

alkaline-earth disilicate glasses. Quantitative estimates of water speciation in the silica glasses agree with those from ²⁹Si NMR⁴. Future proton NMR measurements which resolve different chemical environments will play an important role in elucidating the structures of hydrogen-bearing disordered solids.

The proton magic-angle-spinning (MAS) NMR spectrum for sample SIC, an SiO₂ glass containing 2.5 wt% water quenched from 1,320 °C and 1 kbar (Fig. 1b) consists of an asymmetric peak, broadened towards high frequency. Although this could be explained in terms of a distribution of proton environments, it is known from ²⁹Si NMR data⁴ that this sample contains both hydroxyl groups and molecular water, and we therefore interpret the spectrum in terms of two overlapping resonances. The spectrum can be simulated by two gaussians, a narrow one (full width at half maximum, FWHM = 1.7 p.p.m.) at chemical shift δ = 3.1 p.p.m. and a broader one (FWHM = 5.8 p.p.m.) at δ = 4.2 p.p.m. The narrow peak can be attributed to Si-OH groups, because the proton peak for Suprasil (Fig. 1a) (which is known to contain 1,200 p.p.m. water solely as Si-OH) occurs at δ = 3.3 p.p.m. The peak for molecular water is broader and the side-band envelope more extensive, implying, as expected, a greater degree of H-H dipolar coupling for this species. The ratio of the areas of the Si-OH and molecular-water peaks (including intensity in the side-bands) is 1:2.49 ± 0.30 and, as each water molecule contains two protons, the proportion of dissolved water which reacts with the SiO₂ network according to the reaction Si-O-Si + H₂O = 2Si-OH is therefore 0.45 ± 0.03. This compares with the value of 0.50 ± 0.05 calculated from ²⁹Si data for the same sample⁴.

Sample SIH, an SiO₂ glass containing 8.7 wt% H₂O and synthesized at 1,550 °C and 15 kbar, shows at least four resonances (Fig. 1c). These can be distinguished most clearly in a simulation of the central part of this spectrum (Fig. 1e). The spectrum consists of a broad (FWHM = 5.3 p.p.m.) resonance at δ = 4.5 p.p.m. (gaussian D in the simulation) superimposed on a set of narrower resonances. The very narrow peak at 4.7 p.p.m. (A) resembles that which would result from mobile water present in fluid micro-inclusions within the glass. Isotropically mobile water can be shown to be absent, however, because the static (non-spinning) spectrum does not contain a sharp liquid-like resonance. Furthermore, mobile water would not produce the side-bands observed in the MAS spectrum. The central sharp feature in the MAS spectrum (Fig. 1d) consists of three separate resonances, two sharp ones at 4.7 p.p.m. (A) and 4.3 p.p.m. (B), with FWHMs of 0.26 and 0.80 p.p.m. respectively, and a broader one at 3.2 p.p.m. (C) with a FWHM of 1.4 p.p.m. The two narrowest peaks are better resolved in the first-order spinning side-bands (Fig. 1f), and are also present in the second-order side-bands. Third- and fourth-order spinning side-bands are present for the broadest resonance, but not for the three narrower ones. As the resonances at 3.2 p.p.m. (C) and 4.5 p.p.m. (D) have very similar positions and half-widths to the resonances in SIC, it seems likely that they are due to Si-OH and molecular H₂O, respectively, in environments similar to those in SIC (which has a lower water concentration). The origin of the peaks at 4.7 p.p.m. (A) and 4.3 p.p.m. (B) is more difficult to establish. They are also present in silica glasses with water concentrations between 4.5 and 8.7 wt%, in which the ratio of their intensities appears to be approximately the same as that in SIH. It is possible, therefore, that the resonances are due to two different protons present in the same group. Their relative areas are in the ratio of roughly 1:2, suggesting that peak A could be due to Si-OH and B to H₂O. This assignment is also consistent with the observation that peaks for Si-OH are narrower than those for molecular H₂O. The narrowness of the lines suggests that this group has a very specific geometry. One possible arrangement that might fit these criteria is a water molecule hydrogen-bonded by its oxygen atom to the proton in an Si-OH group (Fig. 2). The possibility of rotation of the H₂O group around its C₂ symmetry axis, leading to a partial averaging

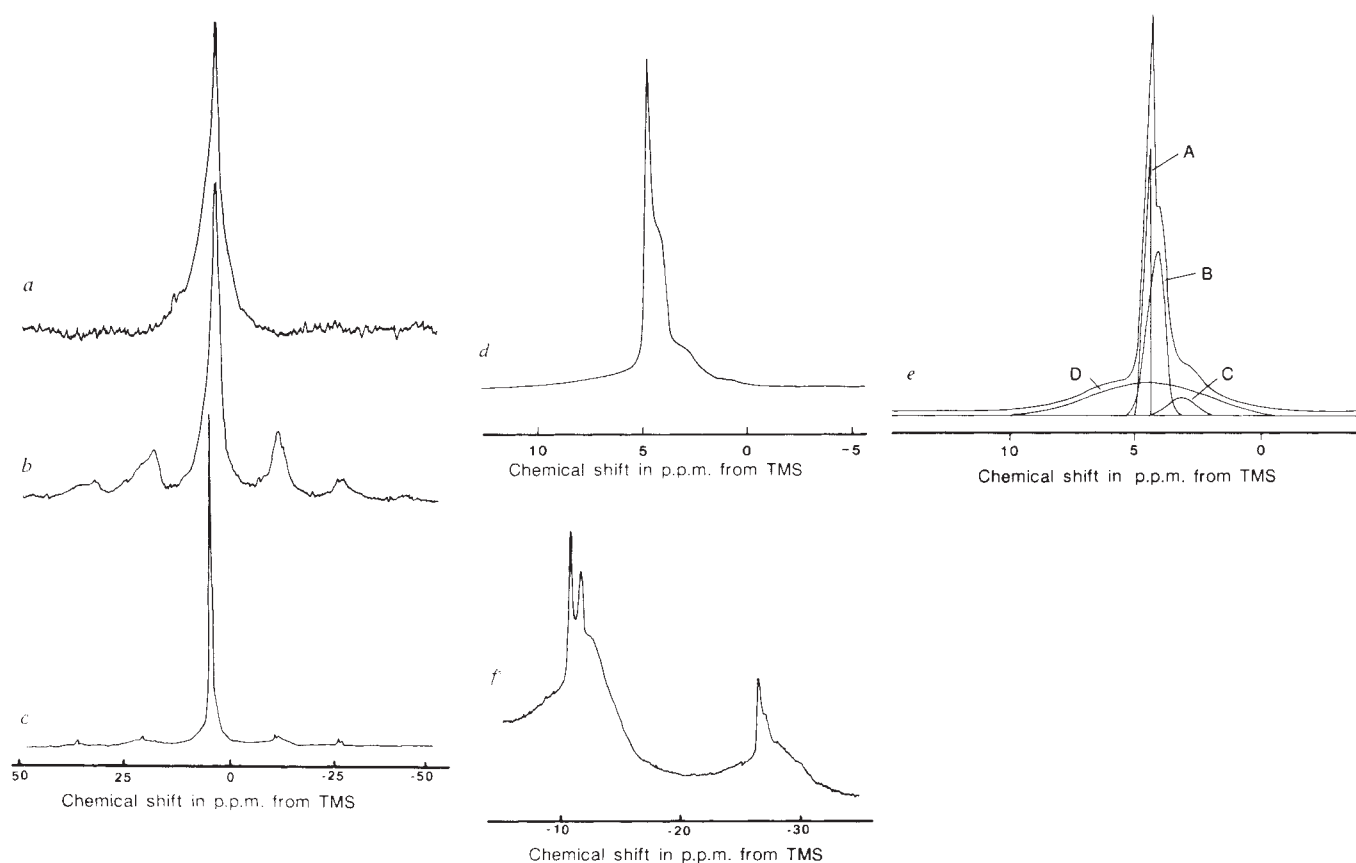


Fig. 1 Proton MAS NMR spectra of hydrous silica glasses. Spectra were obtained using a Bruker MSL 360-MHz spectrometer and a Doty Scientific MAS probe. The reference standard is tetramethylsilane, TMS. Single-pulse experiments were performed on powdered samples spun at 5–6 kHz, using $2\text{-}\mu\text{s}$ ($\sim\pi/2$) pulses and repeating for a few hundred accumulations with recycle delays of 1–60 s, sufficient to prevent saturation. No apodization function was employed. *a*, Suprasil (1,200 p.p.m. (0.12 wt%) H_2O). *b*, Sample SIC, 2.5 wt% H_2O . *c*, *d*, *e* and *f*, Sample SIH, 8.7 wt% H_2O . *d* and *f* have an expanded chemical shift scale, *d* showing the centre band and *f* showing the first-order spinning side-band. *e* is a simulation showing the four components in the spectrum.

of H–H dipolar interactions, could help to explain the narrowness of the proton resonances. Using the simulation of the centre band, and applying a small correction to account for the different relative intensities in the side-bands, we obtain areas (normalized to a total spectrum area of 1) of 0.482, 0.071, 0.156 and 0.291 for the peaks at 4.5(D), 3.2(C), 4.7(A) and 4.3(B) p.p.m. respectively. Given our assignments, the ratio of the total areas for the Si–OH: molecular- H_2O peaks is 0.227:0.773 and implies an extent of reaction of 0.37 ± 0.04 , in agreement with the result obtained from ^{29}Si data⁴.

We have also obtained proton MAS NMR spectra for glasses of albite composition ($\text{NaAlSi}_3\text{O}_8$, with a fully polymerized structure) containing dissolved-water concentrations varying from <1 wt% to ~10 wt%. In all cases the spectra consist of a single broad resonance at 3.5–3.8 p.p.m., similar to those previously reported⁶ but at slightly lower frequency. We were unable, however, to detect differences between the relative intensities of the centre band and the spinning side-bands in these glasses.

The mechanism of water dissolution in de-polymerized melts has been studied by observing proton spectra in three hydrous alkali and alkaline-earth disilicate glasses (Fig. 3). The spectrum for the hydrous sodium disilicate glass (7.8 wt% H_2O , 1,450 °C, 7 kbar) (Fig. 3*a*) consists of two resonances: one at ~4 p.p.m. with an extensive spinning-side-band system, and one at 12 p.p.m. with only two pairs of spinning side-bands. The hydrous barium disilicate glass (3.2 wt% H_2O , 1,350 °C, 10 kbar)

(Fig. 3*b*) has a spectrum similar to that of sodium disilicate except that the two peaks are better resolved. In the hydrous strontium disilicate glass (4.7 wt% H_2O , 1,400 °C, 10 kbar) the two resonances are not completely resolved even at a spinning rate of 8.7 kHz (Fig. 3*c*). However, Fourier transforming from the first rotational echo of the free induction decay shows the presence of a peak at 17 p.p.m. with small spinning side-bands, and a peak at 5 p.p.m. (Fig. 3*d*).

The proton resonances at high frequency (11–17 p.p.m.) indicate the presence of strong hydrogen-bonding. The most likely origin of such bonding is from interactions between Si–OH and

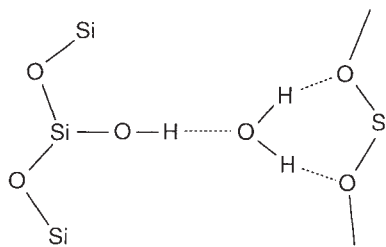


Fig. 2 The group suggested to be responsible for the observed proton peaks in SiO_2 glasses with high concentrations of dissolved water. This group contains molecular water hydrogen-bonded to one Si–OH group and to two bridging oxygen atoms.

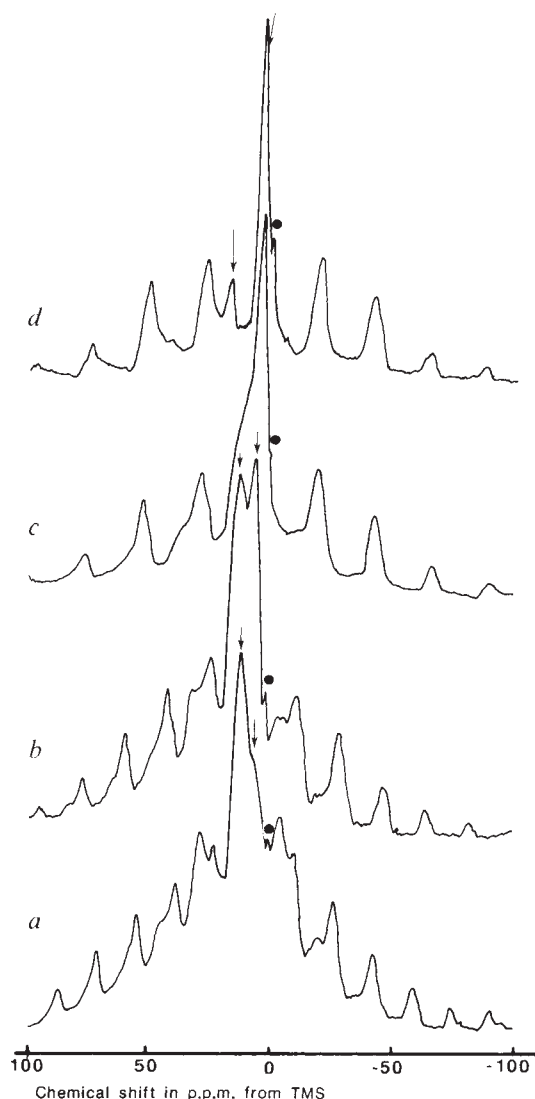


Fig. 3 Proton MAS NMR spectra of hydrous alkali and alkaline-earth disilicate glasses. Accumulation conditions as for Fig. 1, with spinning at 6–10 kHz. *a*, $\text{Na}_2\text{Si}_2\text{O}_5$, 7.8 wt% H_2O . *b*, BaSi_2O_5 , 3.2 wt% H_2O . *c*, SrSi_2O_5 , 4.7 wt% H_2O . *d*, As *c* but Fourier transforming from first rotational echo. Arrows indicate the isotropic chemical shift, and dots indicate a background probe resonance.

O-Si , possibly with both groups attached to the same silicon atom (also suggested on the basis of Raman spectroscopy²). The correlation, observed in ref. 6, between the hydrogen-bonding distance in $\text{O-H}\cdots\text{O}$ and the chemical shift in minerals implies here a very short $\text{O-H}\cdots\text{O}$ distance of ~ 250 pm. The peaks at 4–5 p.p.m. are probably due to molecular water, because they have a large number of spinning side-bands and are at the same position as molecular water in the silica glasses. In all three glasses, the amount of water dissolved in the form of Si-OH appears to be small ($<20\%$).

We acknowledge the financial support of SERC and NERC.

Received 5 October; accepted 22 December 1988.

1. Stolper, E. *Contr. Miner. Petrol.* **81**, 1–17 (1982).
2. McMillan, P. F. & Remmele, R. L. *Am. Miner.* **71**, 772–778 (1986).
3. Mysen, B. O., Virgo, D., Harrison, W. J. & Scarfe, C. M. *Am. Miner.* **65**, 900–914 (1980).
4. Farnan, I., Kohn, S. C. & Dupree, R. *Geochim. cosmochim. Acta* **51**, 2869–2873 (1987).
5. Bartholomew, R. F. & Schreurs, J. W. H. *J. Non-Cryst. Solids* **38/39**, 679–684 (1980).
6. Eckert, H., Yesinowski, J. P., Silver, L. A. & Stolper, E. M. *J. phys. Chem.* **92**, 2055–2064 (1988).

Influence of productivity variations on long-term atmospheric CO_2

Alan C. Mix

College of Oceanography, Oregon State University, Corvallis, Oregon 97331, USA

Evidence from ice cores¹ and deep-sea sediments² shows that atmospheric CO_2 concentration has varied by up to 40% over the past few hundred thousand years. As most of the exchangeable carbon resides in the deep sea, large changes in the atmosphere must have their source here. The distribution of carbon in the ocean is linked to biological productivity, the sinking and degradation of organic matter and calcium carbonate, and ocean circulation³. Carbon-cycle models predict different (and sometimes conflicting) shifts in productivity, and estimates of past productivity constrain the range of possible solutions. Here I use planktonic foraminifera species data in modern and ice-age Atlantic sediments to assess spatial patterns of changes in productivity. Ice-age export productivity was higher than at present by nearly 40% for the whole Atlantic, and by 90% under the Equator. These changes, if extrapolated to the global ocean, support models in which a significant portion of CO_2 changes are driven by variations in biological productivity.

A wide range of models exist to explain ice-age variations in atmospheric p_{CO_2} , each of which has implications for past productivity. Early models call on a $\sim 40\%$ increase in nutrient content of the whole glacial ocean relative to the modern ocean⁴. Such models predict increases in productivity nearly everywhere on the globe (although perhaps mostly in upwelling regions). Past nutrient budgets in the ocean are constrained by cadmium/calcium ratios in benthic foraminifera. Estimates range from a 17% increase in the glacial oceans⁵ to essentially no change⁶. This excludes nutrient budget models as the sole source of changes in CO_2 concentration, and implies that other mechanisms must exist.

A group of high-latitude models consider Antarctic oceanography as a source of CO_2 change. In the Antarctic, CO_2 -rich deep or intermediate waters reach to the sea surface⁷. Either lower exchange of Antarctic surface and deep waters⁸ or higher mixing with warm surface waters⁹ could suppress this polar outcrop of the deep ocean. Both would reduce atmospheric CO_2 . In the southern oceans, unlike the tropics and subtropics, phosphate and nitrate do not limit biological productivity. More efficient use of these nutrients by the Antarctic biota¹⁰ could comprise a biological mechanism to reduce CO_2 . Recent studies suggest that trace-iron is a limiting nutrient in this region¹¹, in which case a higher input of iron-rich dust during the last ice age may have increased productivity and thus reduced atmospheric CO_2 levels¹¹.

All of these high-latitude models permit changes in productivity, but they do not all require it. Comparable CO_2 changes can result either from biological mechanisms or by different modes of circulation independent of the biology. Indeed, in one of the models above⁸, productivity was held constant. Note that model predictions of productivity apply to 'new' production¹², that is, to the net export flux of carbon out of surface waters. It is reasonable to think that enhanced new production might also mean higher primary production, but this is not a strict requirement of the models.

Other models note the effects of changing productivity in the low latitudes. A three-box model¹³ predicts that higher low-latitude upwelling increases productivity and reduces atmospheric CO_2 . This result includes the high latitudes, as upwelling in this model forces overturn of the whole ocean. More complex models include the effect of intermediate water masses. Here equatorial upwelling can operate independently of thermohaline

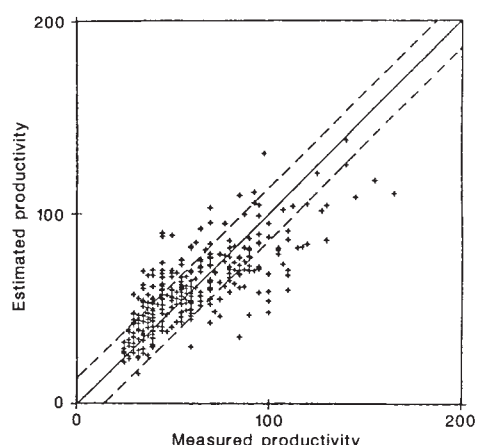


Fig. 1 Calibration of foraminiferal transfer function FAP-6 for primary productivity³⁰. The error envelope (mean absolute value of residuals) (dashed line) is $\pm 12 \text{ g C m}^{-2} \text{ yr}^{-1}$.

overturn. For example, the four-box ocean model of Boyle¹⁴ predicts that doubling the modern shallow upwelling rate in the tropics would produce atmospheric CO_2 by 25 p.p.m. This occurs through transfer of nutrients (and CO_2) from intermediate to deep waters. As with the high-latitude models, there are less nutrients in the polar outcrop of cold water. The upwelling change (by a factor of two) in this model increases the nutrient flux into warm surface waters (and thus increases new productivity) by 36%. By itself this mechanism does not change atmospheric p_{CO_2} as much as observed, but another 25-p.p.m. CO_2 change in this model could come from a 50% reduction of deep-water formation or from lower temperatures¹².

Cadmium¹⁵ and carbon isotope^{16–18} content of benthic foraminifera from the Atlantic support the idea of ice-age nutrient depletion of intermediate waters. There is no consensus, however, as to whether this reflects formation of glacial North Atlantic intermediate water or outflow of Mediterranean water. Carbon isotope data from the Indian¹⁹ and Pacific²⁰ oceans suggest that the ice-age nutrient depletion of intermediate waters may have been global.

Keirs's²¹ fifteen-box model resolves surface, intermediate and deep waters in the Pacific, Atlantic and Indian oceans, as well as polar outcrop areas in both hemispheres. Keir explores a number of options, but favours one in which Antarctic export productivity is two to three times greater than now and the ratio of Antarctic bottom water to North Atlantic deep water is higher than now. This model is similar to Boyle's¹⁴ in that it reduces intermediate water nutrients. It does this by fixing nutrients in the Antarctic, however, and does not require the increase in intermediate-water formation suggested by Boyle. Keir's model decreases warm-ocean productivity by 13 to 43% globally, or by 26 to 50% in the Atlantic sector.

The various models of the carbon system discussed above imply different regional changes in productivity. For the high latitudes, results range from no change to an increase by a factor of three. In the low latitudes anything from a factor-of-two increase to a 50% decrease at the glacial maximum is possible.

An obvious test of these model predictions is to reconstruct past productivity.

Initial attempts at reconstructing productivity have used organic-carbon data²². This is a logical choice, as organic-carbon accumulation rates in the deep-sea sediments should in some way relate to the export flux of organic carbon from the sea surface. A recent summary of organic-carbon data suggests that higher productivity occurred in many places at the last glacial maximum²³. This inference, although useful, has not met universal acceptance. Possible causes of poor signal preservation include the diverse composition of bulk organic matter²⁴, the effects of deep-sea oxygen²⁵, sensitivity to errors in sedimentation rates²⁶ and contamination with organic matter from the continents²⁷.

Because of these uncertainties regarding organic carbon, I have estimated past productivity in a different way. A transfer function²⁸ may be used to relate the modern planktonic foraminifera found in deep-sea sediments²⁹ to modern productivity. The equation coefficients are constants, found by multiple regression of the core-top foraminiferal assemblages on the modern productivity values from each site.

To estimate past productivity, down-core species abundances are first reduced to factor assemblages as defined by the core tops; these are operated on the productivity transfer function. Both the assemblages and the productivity equation are similar in form to the transfer functions used to estimate sea surface temperatures from the same foraminiferal data. Both temperature and productivity may be estimated from the same data set if the two parameters are statistically independent²⁸, which is the case here. In the modern world, productivity and mean sea surface temperature are not correlated on a large scale. For the 356 calibration sites used here, the correlation between productivity and sea surface temperature is essentially zero ($r = -0.11$)³⁰. Thus, the productivity transfer function, if designed properly, is independent of variations caused by temperature changes.

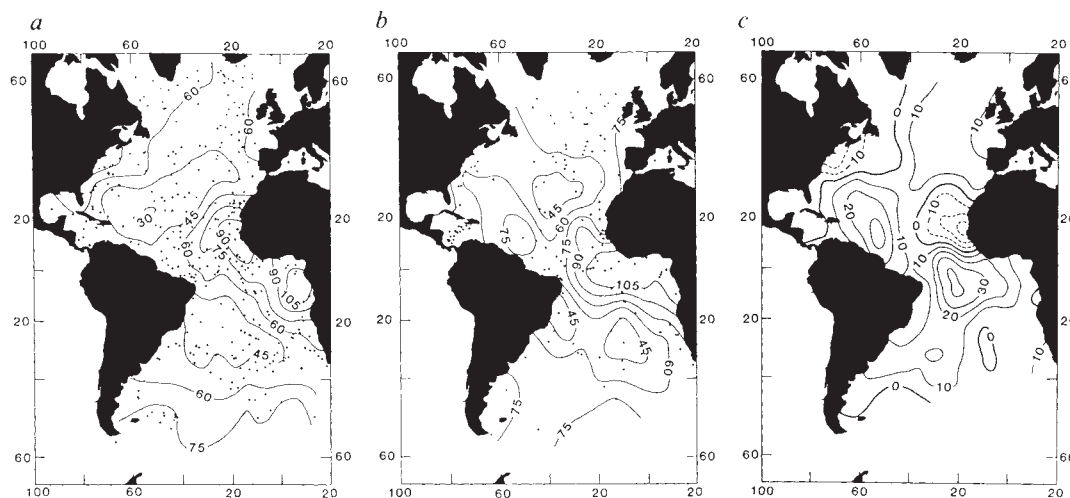
Calibration of the transfer function requires a map of modern productivity. Berger *et al.*³¹ have compiled primary-productivity data based on ^{14}C incubation. They note the uncertainties involved with this method, but assert that the large-scale patterns are reasonably reliable. Using these data and a model based on nutrient concentrations, light intensity and continental influence, they develop a global map of productivity. Although some assumptions were made to derive this map, I use it here for calibration because it is the only available map that has global coverage and uniform smoothing. It is very similar to earlier global productivity maps³², but includes much more recent data. Further studies will no doubt expand on the modern data set, and this will improve future attempts at calibrating the geological record.

Transfer function FAP-6 (Foraminiferal Atlantic Productivity equation Six) is used here (Fig. 1). Full documentation of the factor and equation coefficients appears elsewhere³⁰. The function has an error envelope of $\pm 12 \text{ g carbon m}^{-2} \text{ yr}^{-1}$ (the average absolute value of residuals) and reconstructs more than 61% of the modern variance in productivity. This compares with the error envelope of $\pm 30 \text{ g C m}^{-2} \text{ yr}^{-1}$ for primary-productivity estimates based on organic-carbon content of sediments²³. At sites with productivity $\leq 100 \text{ g C m}^{-2} \text{ yr}^{-1}$, the calibration errors of

Table 1 Regional changes in productivity between the ice age and present

Area	Grid points	Primary productivity ($\text{g C m}^{-2} \text{ yr}^{-1}$)			New productivity ($\text{g C m}^{-2} \text{ yr}^{-1}$)		
		Modern	18 kyr	% change	Modern	18 kyr	% change
Equatorial	40	70	96	+37	12.3	23.0	+87
Subtropical	193	47	57	+21	5.5	8.1	+47
Subpolar	82	62	67	+8	9.6	11.2	+17
Eastern boundary	51	74	75	+1	13.7	14.1	+3
Total	366	57	67	+18	8.1	11.2	+38

Fig. 2 Atlantic primary productivity, in $\text{g C m}^{-2} \text{ yr}^{-1}$, using equation FAP-6 and CLIMAP²⁹ foraminiferal species data. *a*, Core tops (modern). *b*, Last glacial maximum (~18 kyr BP). *c*, The productivity difference (glacial minus modern). Positive values indicate higher ice-age productivity.



the two techniques are more similar: errors of the foraminiferal equation are $\pm 11 \text{ g C m}^{-2} \text{ yr}^{-1}$, whereas those of the organic-carbon equation are $\pm 19 \text{ g C m}^{-2} \text{ yr}^{-1}$. Errors of the foraminiferal estimates at productivities $> 100 \text{ g C m}^{-2} \text{ yr}^{-1}$ are $\pm 28 \text{ g C m}^{-2} \text{ yr}^{-1}$, compared with $\pm 53 \text{ g C m}^{-2} \text{ yr}^{-1}$ for the organic-carbon equation.

A low bias in the estimates of the foraminiferal equation may exist in the highest-productivity areas (Fig. 1). This suggests that after a certain point further increases of productivity do not modify the faunal assemblages. Another bias in the foraminiferal productivity estimates could come from selective dissolution of carbonate shells, but this effect is not detected for moderate dissolution. There is no relationship between core-top residuals and water depth, a good predictor of dissolution intensity, over a range from 200 to 5,500 m (ref. 30).

Primary productivity of the Atlantic Ocean is calculated from core-top and glacial-maximum sediments (18 kyr BP) using the foraminiferal-species data of CLIMAP²⁹. At present, productivity is low ($< 30 \text{ g C m}^{-2} \text{ yr}^{-1}$) in the central subtropical gyres (Fig. 2*a*), and is highest ($> 100 \text{ g C m}^{-2} \text{ yr}^{-1}$) in the eastern boundaries (Canary and Benguela currents). At the last glacial maximum (Fig. 2*b*), the central gyres continued to have the lowest productivity, $\sim 50 \text{ g C m}^{-2} \text{ yr}^{-1}$, which is slightly higher than at present. Highest ice-age values ($\sim 120 \text{ g C m}^{-2} \text{ yr}^{-1}$) occur along the Equator rather than in the eastern boundary currents.

Differences between the glacial and modern maps are shown in Fig. 2*c*. The largest increase in ice-age productivity occurs under the South Equatorial current. This may be compared with organic-carbon data from the area, which indicate a doubling of mass accumulation rates at the glacial maximum³³. Ice-age productivity is lower than that at present off north-west Africa and along the North American margin (Fig. 2*c*). The result off north-west Africa conflicts with a result based on organic carbon²³, which infers increased productivity in this region at the glacial maximum, but it may be consistent with lower glacial carbonate fluxes (thought to reflect lower productivity) on the Sierra Leone Rise³⁴. Because the sites off north-west Africa which were studied for organic carbon are mostly under coastal upwelling, whereas the cores used in this work are mostly offshore, the two data sets are not easily comparable. It is possible that the histories of coastal and offshore productivity are different.

High productivity at the Equator is consistent with ice-age cooling^{29,35}. Upwelling could drive both effects. Cooling also occurred off north-west Africa^{29,35}, however, where the foraminiferal data suggest lower productivity. Cooling here may reflect advection of cool surface waters from higher latitudes instead of upwelling of cold nutrient-rich sub-surface waters. Again, note that the cores used here are not under the coastal

upwelling system, but are offshore. Ice-age productivity appears to have increased not only in the equatorial upwelling area but also in the subtropics (Fig. 2*c*), so equatorial upwelling is not the only mechanism contributing to the change. Although cadmium data preclude large variations in nutrient budgets, small changes remain possible^{5,6}. In addition, stronger ice-age winds may have enhanced vertical mixing outside of the upwelling zones.

Table 1 lists changes in productivity by area. The equatorial area has boundaries of 60° W , 0° W , 8° S , 8° N . The other areas have parts in both the Northern and Southern Hemispheres. The subtropical boundaries are 90° W , 25° W , 8° N , 40° N and 80° W , 0° E , 40° S , 8° S . For the subpolar regions the boundaries are 80° W , 0° E , 40° N , 65° N and 70° W , 20° E , 60° S , 40° S . The eastern boundary current (EBC) regions are defined as 25° W , 0° E , 8° N , 40° N and 0° E , 20° E , 40° S , 0° S . The major increases in primary productivity at the last glacial maximum occur in the equatorial and subtropical areas, with increases of 37% and 21% respectively. The subpolar and EBC regions change little on average. For the EBC this reflects a slight increase in productivity in the Southern Hemisphere and a slight decrease in the Northern Hemisphere. Over the whole Atlantic, the mean glacial primary productivity ($67 \pm 1.6 \text{ g C m}^{-2} \text{ yr}^{-1}$) was $18 \pm 5\%$ higher than at present ($57 \pm 1.0 \text{ g C m}^{-2} \text{ yr}^{-1}$).

Only the portion of the organic matter exported to the deep sea, or 'new' productivity, drives partitioning of CO_2 in the ocean. We cannot yet constrain the recycling of carbon within near-surface waters in the past, but modern primary and new productivity may be related. If we use the proposed modern relationship¹², the $18 \pm 5\%$ glacial-interglacial change in primary productivity amplifies to a $38 \pm 12\%$ increase in new productivity. For the equatorial band, glacial new productivity is nearly twice that at present (Table 1). The transform between primary and new productivity is uncertain even for the present ocean, and is only an approximation. Some conclusions can be drawn, however. The changes suggested here agree with models in which atmospheric CO_2 is driven lower by increasing low-latitude productivity¹⁴. They are not consistent with models that require lower warm-ocean productivity at the last glacial maximum²¹.

This study covers the Atlantic ocean alone, which accounts for about one-third of the present total oceanic production (both primary and new productivity)¹². If productivity in the Pacific and Indian oceans varied in a different fashion to that in the Atlantic, the inferences made here could be in error. A recent global study of organic carbon reconstructs an ice-age increase in new productivity of $45 \pm 6\%$ over the modern value²³. Although the spatial patterns disagree in detail with the estimates made here, the average change is remarkably consistent with the $38 \pm 12\%$ increase suggested by the foraminiferal data. To

further understand the mechanisms responsible for these changes, we must work toward global reconstructions with regional resolution. From the reconstruction presented here, it appears that equatorial upwelling and vertical mixing in the subtropics are two possible mechanisms.

I appreciate assistance from C. Peterson on foraminiferal species counts and data analysis. Discussions with W. Berger, S. Emerson, M. Sarnthein and E. Suess, and a review by L. Labeyrie added to this paper. This research and curation of cores was supported by the NSF.

Note added in proof: A new model³⁶ explores several mechanisms for lower glacial p_{CO_2} and produces a range of effects on nutrient fluxes to the warm surface ocean (and thus on new productivity). These vary from an ~20% decrease (in cases without enhanced upwelling) to an ~120% increase (for a case with higher glacial upwelling) relative to modern values. The reconstruction presented here is most similar to, but less extreme than, the cases with enhanced tropical upwelling.

Received 11 July; accepted 1 December 1988.

1. Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **329**, 408–414 (1987).
2. Shackleton, N. J., Hall, M. A., Line, J. & Shu, C. *Nature* **306**, 319–322 (1983).
3. Broecker, W. S. & Peng, T.-H. *Radiocarbon* **28**, 309–327 (1986).
4. Broecker, W. S. *Geochim. cosmochim. Acta* **46**, 1689–1705 (1982).
5. Boyle, E. A. *Geochim. cosmochim. Acta* **50**, 265–276 (1986).
6. Boyle, E. A. *Paleoceanography* **3**, 471–490 (1988).
7. Takahashi, T., Broecker, W. S. & Bainbridge, A. E. in *Carbon Cycle Modeling* (ed. Bolin, B.) 159–199 (Wiley, New York, 1981).
8. Toggweiler, J. R. & Sarmiento, J. L. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archaean to Present* (eds Sundquist, E. T. & Broecker, W. S.) Monogr. 32, 163–184 (Am. Geophys. Un., Washington DC, 1985).
9. Ennever, F. K. & McElroy, M. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations*

- Archaean to Present* (eds Sundquist, E. T. & Broecker, W. S.) Monogr. 32, 154–162 (Am. Geophys. Un., Washington DC, 1985).
10. Wenk, T. & Siegenthaler, U. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archaean to Present* (eds Sundquist, E. T. & Broecker, W. S.) Monogr. 32, 185–194 (Am. Geophys. Un., Washington DC, 1985).
11. Martin, J. H. & Fitzwater, S. E. *Nature* **331**, 341–343 (1988).
12. Eppley, R. W. & Peterson, B. J. *Nature* **282**, 677–680 (1979).
13. Siegenthaler, U. & Wenk, T. *Nature* **308**, 624–626 (1984).
14. Boyle, E. A. in *Mesozoic and Cenozoic Oceans* (ed. Hsu, K. J.) 337–351 (Am. Geophys. Un., Washington DC, 1986).
15. Boyle, E. A. & Keigwin, L. D. *Nature* **330**, 35–40 (1987).
16. Cofer-Shabica, N. & Peterson, L. *Geol. Soc. Am. Abstr. Prog.* **18**, 567 (1986).
17. Oppo, D. & Fairbanks, R. G. *Earth planet. Sci. Lett.* **86**, 1–15 (1987).
18. Zahn, R., Sarnthein, M. & Erlenkeuser, H. *Paleoceanography* **2**, 543–560 (1987).
19. Kallel, N., Labeyrie, L. D., Juillet-Leclerc, A. & Duplessy, J.-C. *Nature* **333**, 651–655 (1988).
20. Duplessy, J.-C. *et al. Paleoceanography* **3**, 343–360 (1988).
21. Keir, R. *Paleoceanography* **3**, 413–446 (1988).
22. Muller, P. J. & Suess, E. *Deep-Sea Res.* **26A**, 1347–1362 (1979).
23. Sarnthein, M., Winn, K., Duplessy, J.-C. & Fontugne, M. R. *Paleoceanography* **3**, 361–399 (1988).
24. Suess, E. & Muller, P. J. *Colloq. Intern. CNRS* **293**, 17–26 (Ed. Centr. Natn. Rech. Sci., Paris, 1980).
25. Emerson, S. & Hedges, J. *Paleoceanography* **3**, 621–634 (1988).
26. Muller, P. J., Erlenkeuser, H. & von Grafenstein, R. in *Coastal Upwelling, its Sediment Record, Part B: Sedimentary Record of Ancient Coastal Upwelling* (eds Thiede, J. & Suess, E.) 365–398 (Plenum, New York, 1983).
27. Prah, F. & Muehlhausen, L. in *Productivity of the Ocean: Present and Past* (eds Berger, W. H., Smetacek, V. S. & Wefer, G.) 271–289 (Wiley, New York, 1989).
28. Imbrie, J. & Kipp, N. G. in *The Late Cenozoic Glacial Ages* (ed. Turekian, K. K.) 77–181 (Yale Univ. Press, New Haven, 1971).
29. CLIMAP GSA Map and Chart Series MC-36 (1981).
30. Mix, A. C. in *Productivity of the Ocean: Present and Past* (eds Berger, W. H., Smetacek, V. S., & Wefer, G.) 313–340 (Wiley, New York, 1989).
31. Berger, W. H., Fischer, K., Lai, C. & Wu, G. *Scripps Inst. Oceanogr. Ref.* 87–30 1–67 (1987).
32. Koblenz-Mischke, O. I., Volkovinskiy, V. V. & Kabanova, J. G. in *Scientific Exploration of the South Pacific* (ed. Wooster, W.) 183–193 (Natl. Acad., Washington DC, 1970).
33. Lyle, M., *Nature* **335**, 529–532 (1988).
34. Curry, W. & Lohmann, G. P. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archaean to Present* (eds Sundquist, E. T. & Broecker, W. S.) Monogr. 32, 285–301 (Am. Geophys. Un., Washington DC, 1985).
35. Mix, A. C., Ruddiman, W. F. & McIntyre, A. *Paleoceanography* **1**, 43–66 (1986).
36. Boyle, E. A. *J. geophys. Res.* **93**, 15701–15714 (1988).

Changes in copper, zinc and cadmium concentration in Antarctic ice during the past 40,000 years

Françoise Batifol, Claude Boutron* & Martine de Angelis

Laboratoire de Glaciologie et Géophysique de l'Environnement du CNRS, Domaine Universitaire, BP 96, 38402 St Martin D'Heres Cedex, France

The occurrence of toxic metals^{1–3} in ancient Antarctic ice provides an excellent means of assessing the past natural tropospheric cycles of these metals, knowledge of which is vital for modern time trends to be judged in proper perspective^{4–6}. In spite of this, there are at present very few reliable data for metals in Antarctic ice, with the exception of lead^{7,8}. Here we present comprehensive data on the time variations of copper, zinc and cadmium in Antarctic ice from 39,500 to 2,700 yr BP. During this time, the concentrations of these metals have varied significantly, the highest values being observed during the Last Glacial Maximum (LGM). Soil dust appears to have been the main natural source of metals during the LGM, but during the Holocene, it comprises only part of the measured concentrations. The Holocene excesses might be explained by a volcanic or biogenic contribution.

We have analysed 38 sections of the 905-m Antarctic Dome C ice core using ultraclean laboratory procedures. This core was thermally drilled in a dry bore hole at Dome C (77°39'S, 124°10'E, elevation 3,240 m above sea level) during the 1977–78 summer field season, and has been shown to cover the past 40,000 yr (refs 9, 10). Each section (10.5 cm in diameter, 15–20 cm in length) was cleaned by a rinsing-melting technique inside the Grenoble clean laboratory. The core was

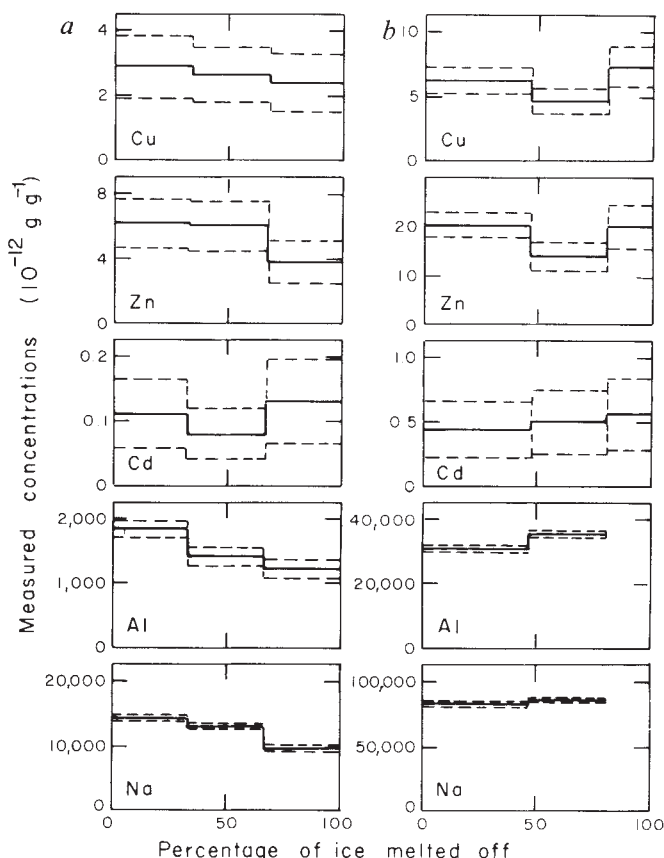
rinsed twice with flowing ultrapure water until about 30% of the original sample is melted off. The core was then placed inside an ultraclean wide-mouth polyethylene bottle for three successive rinsings with ultrapure water. The remaining core (100–300 g, that is, about 30% of the original core) was then melted inside the polyethylene bottle, and pre-concentrated (by a factor of 30) by non-boiling evaporation in pre-conditioned FEP teflon beakers inside the Grenoble clean laboratory. The sample was then analysed for Cu, Zn and Cd by graphite-furnace atomic-absorption spectrometry. Blanks for the successive steps of the analytical procedure were carefully determined, as described in detail elsewhere¹¹. The overall uncertainty of the data is mainly dependent on blank variability. For Cu, it ranges from ~15% for the high concentrations observed during the LGM to ~80% for the very low concentrations observed during the Holocene. For Zn and Cd, the corresponding ranges are 10–40% and 50–100%, respectively¹¹. We also analysed all of the samples for sodium and aluminium (by instrumental neutron activation) and some of them for sulphate (by ion chromatography), both without pre-concentration.

To test the efficiency of the decontamination procedure, we investigated the variations in concentration of the various elements from the outside to the inside of four decontaminated sections (two from the Holocene, one from the LGM and one from the pre-LGM glacial period). These variations were determined by melting each of these decontaminated sections in several successive steps. The first melted fraction corresponded to the outside of the decontaminated core, and the last one to the innermost part. Each fraction was analysed separately. Concentrations in these successive fractions were found to be fairly stable for all the measured elements (Fig. 1); in most cases, the observed fluctuations are within the experimental precision of the data.

Figure 2 shows the variations of Cu, Zn and Cd concentrations as a function of the depth and age of the ice during the past 40,000 yr. The variation profiles are complete only for Cu and Zn; for Cd, the measured concentrations provide upper limits only (<0.1 pg Cd g⁻¹ or <0.3 pg Cd g⁻¹) for 24 of the 38 sections.

* To whom correspondence should be addressed. Also at UFR de Physique, Université de Grenoble, Domaine Universitaire, BP 68, 38402 St Martin D'Heres Cedex, France.

Fig. 1 Measured Cu, Zn, Cd, Al and Na concentrations in successive melted fractions of two decontaminated ice sections (a, 241.1 m depth (5,960 yr BP); b, 738.8 m depth (29,800 yr BP)). Concentrations are expressed as a function of the percentage of the decontaminated sample melted off: 0% corresponds to the outside of the decontaminated core, and 100% to the centre. Dotted lines show the estimated analytical uncertainties.



(The upper limit is not the same for all the 24 samples for which Cd concentration is given as upper limit only.) Table 1 shows mean metal concentrations for each of the four climatic stages covered by the Dome C core. For Cu and Cd, our data are the first comprehensive ones to be published for ancient Antarctic ice. The only data published previously for these two metals were for a single block of blue ice more than 12,000 yr old from coastal East Antarctica¹², for which the reported value of $17.5 \text{ pg Cu g}^{-1}$ is consistent with our data. On the other hand, the value of 2.6 pg Cd g^{-1} reported in ref. 12 is higher than any of our results, and has been questioned elsewhere^{4,14}. For Zn, data has been published both in ref. 12 and for 41 sections of the Dome C core¹³. For the Holocene, our values are lower than those of this last study, probably because of the less accurate analytical procedures used in the earlier study.

As shown in Fig. 2, Cu and Zn concentrations in Antarctic ice have varied significantly during the past 40,000 yr. They were rather low from around 39,500 to 30,000 yr BP (means of 5 pg Cu g^{-1} and 11 pg Zn g^{-1}); they rose rapidly to high values and remained thus (means of 18 pg Cu g^{-1} and 44 pg Zn g^{-1}) during the last part (LGM) of the last ice age (from ~25,000 to 16,000 yr BP); they then declined rapidly to very low values during the LGM-Holocene transition, and remained very low (means of 2.7 pg Cu g^{-1} and 5.7 pg Zn g^{-1}) during the Holocene. For Cd, the variation profile is not so clear because most of the concentration values are available as upper limits only. There is a suggestion, however, that Cd concentrations could have been relatively high during parts of the last ice age.

Aluminium concentrations (Fig. 2) provide an index to calculate the contributions to metal concentrations from wind-blown soil dust, using the mean metal/aluminium ratios in soil¹⁵, these being 4.2×10^{-4} for Cu, 1.27×10^{-3} for Zn and 4.9×10^{-6} for Cd. We have used soil ratios rather than crustal ratios because wind-blown dust transported to Antarctica from other continents is likely to originate mainly from soil. For Cu and Zn, the calculated soil-dust contribution is close to the measured concentrations in the ice from about 39,500 to 13,500 yr BP (Fig. 2), that is, during the pre-LGM part of the last ice age, the LGM, and the first half of the LGM-Holocene transition. This

establishes that during that time interval, natural Cu and Zn in the global troposphere originated mainly from wind-blown soil dust, as already found for Pb^{7,8}. During the second half of the LGM-Holocene transition and during the Holocene (since about 13,500 yr BP), the dust contribution comprises only a fraction of the measured Cu and Zn concentrations (about 15% and 25% respectively). It appears, therefore, that since about 13,500 BP there has existed a significant natural excess of Cu and Zn above that contributed by wind-blown soil dust. The situation is again less clear for Cd as part of our data indicate upper limits only. For part of our LGM and pre-LGM samples, however, the soil-dust contribution is found to represent a large fraction of the measured Cd concentrations. This suggests that, at least during the end of the last ice age, a large fraction of Cd

Table 1 Measured concentrations in the Antarctic Dome C ice core

Climatic stage	Measured concentrations*					
	Cu (pg g^{-1})	Zn (pg g^{-1})	Cd (pg g^{-1})†	Al (ng g^{-1})	Na (ng g^{-1})	SO ₄ (ng g^{-1})‡
Holocene§	2.7 ± 1.0 (1.4–4.3)	5.7 ± 3.4 (1.8–13.6)	— (<0.1 –0.66)	1.08 ± 0.56 (0.41–2.1)	17 ± 4 (13–24)	72 ± 9 (56–85)
Holocene–last ice age transition	2.8 ± 1.2 (1.3–5.2)	9.2 ± 9.3 (1.1–26)	— (<0.1 –0.32)	3.6 ± 3.0 (0.56–10)	31 ± 11 (13–51)	91 ± 14 (65–113)
Last Glacial Maximum¶	18 ± 9 (6.4–33)	44 ± 19 (15–62)	— (<0.1 –1.00)	64 ± 21 (27–92)	99 ± 21 (71–128)	172 ± 20 (160–190)
Pre-LGM part of the last ice age#	5.0 ± 1.6 (2.9–6.2)	11 ± 3.7 (7.6–17)	— (<0.3 –0.51)	15 ± 8.7 (6.9–31)	69 ± 27 (20–91)	156 ± 19 (137–174)

For each element and each climatic stage, the first line gives the mean measured concentration and the corresponding standard deviation, and the second line gives the minimum and the maximum measured concentrations. The complete data set is given in ref. 11.

* Number of ice sections analysed: 11 for the Holocene, 10 for the Holocene–last ice age transition, 11 for the LGM, and 6 for the pre-LGM part of the last ice age.

† For Cd, no mean concentrations and no standard deviations are given as many concentration values were obtained as upper limits only.

‡ For SO₄, the number of sections analysed for each of the four successive climatic stages were 7, 8, 6 and 3, respectively.

§ Depth (real depths) range: 127–403 m (2,700–10,700 yr BP).

|| Depth range: 411–499 m (11,000–14,300 yr BP).

¶ Depth range: 546–671 m (17,000–25,200 yr BP).

Depth range: 739–878 m (29,800–39,500 yr BP).

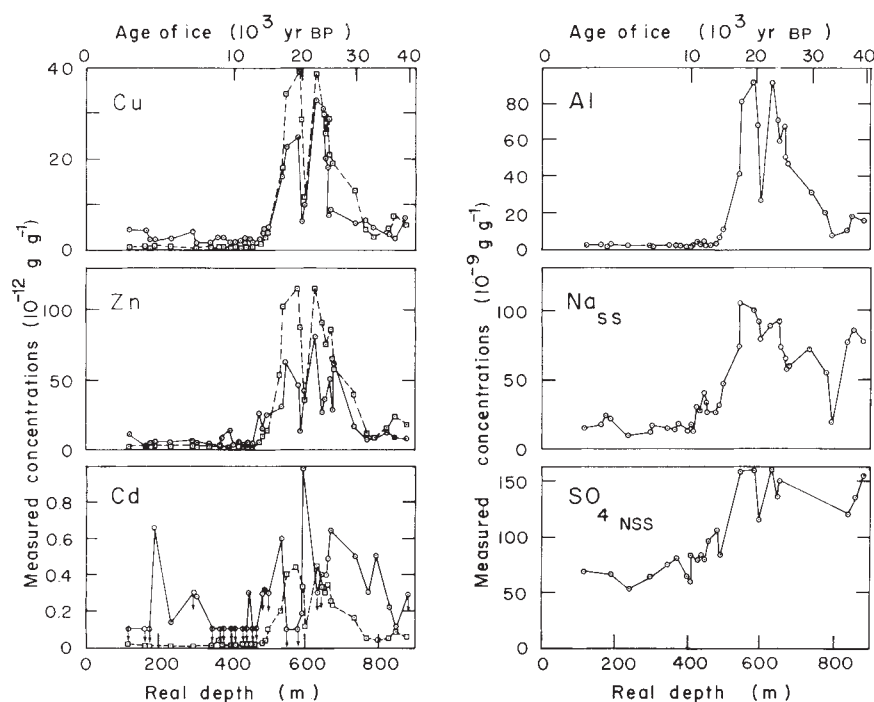


Fig. 2 Variations of Cu, Zn, Cd, Al, seasalt Na and non-seasalt SO_4 concentrations in the Dome C ice core (open circles). For Cu, Zn and Cd, variations of the soil-dust contribution calculated from the mean metal/aluminium ratios in soil¹⁵ are also shown (open squares). Depths are given as real depths rather than as metres of ice equivalent. Approximate ages shown at the top of the figure are from refs 9 and 10.

in the global troposphere probably originated, like Cu and Zn, from wind-blown soil dust.

After correction for contribution from dust, the measured sodium concentrations (Fig. 2) provide an index for estimating the contribution to metal concentrations from the oceans. This contribution can be estimated from the mean metal/sodium ratios in present-day bulk sea water (5.5×10^{-9} for Cu (ref. 16), 4.6×10^{-10} for Zn (ref. 17) and 1.8×10^{-11} for Cd (ref. 18)) combined with the very few published enrichment factors relative to bulk sea water for these metals in present-day marine aerosols (1,000 for Cu, 50,000 for Zn and 10,000 for Cd (refs 18–20)). It must be noted that both the metal/sodium ratios and the metal enrichment factors were probably lower for the unpolluted pristine ocean several thousand years ago than they are now. The oceanic contribution is found to be very small both during the last ice age and during the Holocene. This confirms that the oceans are not an important source of natural Cu, Zn and Cd to the global troposphere, as found already for Pb (refs 7, 8).

A tentative estimate of the contribution to metal concentrations from volcanoes can be made from the non-seasalt sulphate concentrations measured in our samples (Fig. 2) and mean metal/sulphur ratios in volcanic emissions. We have assumed that a constant percentage of non-seasalt sulphate in the ice was of tropospheric volcanic origin and that our samples did not include any stratospheric volcanic sulphate. From global inventories of natural sulphur sources and marine data^{7,21,22}, we estimate that this percentage was $\sim 13\%$ during the Holocene and that it was not significantly different during the LGM. From available data, it is difficult to estimate the mean metal/sulphur ratios in volcanic emissions. However, we have made tentative estimates of these ratios from the mean Pb/S ratio, 2.5×10^{-5} , in volcanic emissions calculated by Patterson and Settle²³, and from the best available global volcanic Pb, Cu, Zn and Cd fluxes available in the literature^{18,24–26}. Mean estimated ratios are 5×10^{-4} for Cu/S and Zn/S, and 5×10^{-5} for Cd/S. The volcanic contribution is found to range from about 0.5 to 5 pg g^{-1} for Cu and Zn, and from about 0.07 to 0.4 pg g^{-1} for Cd. When compared with the observed excesses of metal above that contributed by soil dust during the Holocene, these estimates (which involve rather large uncertainties) suggest that the volcanic contribution could account for a large part of these excesses. They also confirm that during the LGM, this volcanic contribu-

tion was much smaller than the soil dust contribution, at least for Cu and Zn.

Recent inventories²⁵ of natural sources of metals to the global troposphere have emphasized that biogenic sources might be very significant. From our data, it is impossible to make any estimate of these biogenic contributions, but it seems possible that they could explain a significant part of the excesses of metal, above that contributed by soil dust, observed during the Holocene.

This work was supported by the Ministère de l'Environnement, the Expéditions Polaires Françaises, the Terres Australes et Antarctiques Françaises and the Division of Polar Programs of the NSF. We thank C. Feniet for performing the ion-chromatography measurements, and R. Delmas, M. Legrand and P. Buat-Menard for many constructive criticisms.

Received 25 October; accepted 29 December 1988.

- Wood, J. M. in *Global Chemical Cycles and their Alterations* (eds Stumm, W. & Pytkowicz, R. M.) 137–153 (Dahlem Konferenzen, Berlin, 1976).
- Craig, P. J. in *The Natural Environment and Biogeochemical Cycles* (ed. Hutzinger, O.) 169–227 (Springer, Berlin, 1980).
- Nriagu, J. O. & Pacyna, J. M. *Nature* **333**, 134–139 (1988).
- Wolff, E. W. & Peel, D. A. *Nature* **313**, 535–540 (1985).
- Boutron, C. in *Heavy Metals Cycling* (eds Hutchinson, T. C. & Subramanian, V.) Scope Rep. 31, (Wiley, New York, in the press).
- Peel, D. A. in *The Environmental Record in Glaciers*, Dahlem Konferenzen (eds Langway, C. C. Jr & Oeschger, H.) 207–223 (Wiley, Chichester, 1989).
- Boutron, C. & Patterson, C. C. *Nature* **323**, 222–225 (1986).
- Boutron, C., Patterson, C. C., Petrov, V. N. & Barkov, N. I. *Atmos. Environ.* **21**, 1197–1202 (1987).
- Jouzel, J. *et al.* *Quat. Res.* (in the press).
- Benoist, J. P. thesis, Univ. Grenoble (1986).
- Batifol, F. thesis, Univ. Grenoble (1987).
- Boutron, C., Leclerc, M. & Risler, N. *Atmos. Environ.* **18**, 1947–1954 (1984).
- Petit, J. R., Briat, M. & Royer, A. *Nature* **293**, 391–394 (1981).
- Wolff, E. W. & Peel, D. A. *Ann. Glaciol.* **7**, 61–69 (1985).
- Bowen, H. J. M. *Environmental Chemistry of the Elements* (Academic, New York, 1979).
- Boyle, E. A., Chapnick, S. D., Shen, G. T. & Bacon, M. P. *J. geophys. Res.* **91**, 8573–8593 (1986).
- Bruland, K. W. *Earth planet. Sci. Lett.* **47**, 176–198 (1980).
- Duce, R. A., Arimoto, R., Ray, B. J., Unni, C. K. & Harder, P. J. *J. geophys. Res.* **88**, 5321–5342 (1983).
- Arimoto, R., Duce, R. A., Ray, B. J., Hewitt, A. D. & Williams, J. J. *J. geophys. Res.* **92**, 8465–8479 (1987).
- Weisel, C. P., Duce, R. A., Fashing, J. L. & Heaton, R. W. *J. geophys. Res.* **89**, 11607–11618 (1984).
- Zehnder, A. & Zinder, S. in *The Handbook of Environmental Chemistry* (ed. Hutzinger O.) Vol. 1, Part A, 105–145 (Springer, New York, 1980).
- Berresheim, H. & Jaeschke, W. *J. geophys. Res.* **88**, 3732–3740 (1983).
- Patterson, C. C. & Settle, D. M. *Geochim. cosmochim. Acta* **51**, 675–681 (1987).
- Nriagu, J. O. *Nature* **279**, 409–411 (1979).
- Nriagu, J. O. in *Control and Fate of Atmospheric Heavy Metals* (eds Pacyna J. M. & Ottar B.) (in the press).
- Lambert, G., Le Cloarec, M. F. & Pennisi, M. *Geochim. cosmochim. Acta* **52**, 39–42 (1988).

Effects of sedimentary sorting on neodymium isotopes in deep-sea turbidites

S. M. McLennan*†, M. T. McCulloch*, S. R. Taylor* & J. B. Maynard‡

* Research School of Earth Sciences, Australian National University, Canberra, ACT, Australia

† Department of Earth and Space Sciences, State University of New York at Stony Brook, Stony Brook, New York 11794-2100, USA

‡ H. N. Fisk Laboratory for Sedimentology, Department of Geology, University of Cincinnati, Cincinnati, Ohio 45221, USA

The study of neodymium model ages of sedimentary rocks has added considerably to our understanding of crustal growth and evolution. So far, however, the possible effects of various sedimentary processes on Nd isotope distributions have not been fully evaluated. Here we present Nd isotope data for a series of sand-mud pairs in modern deep-sea turbidites from different tectonic settings, which show that sands and muds commonly have significantly different Nd isotope compositions that in only some cases are associated with differences in Sm/Nd ratio. Differences are as great as $7\epsilon_{Nd}$ units, corresponding to Nd model ages that differ by up to 0.44 Gyr, and may be explained by separation (or unmixing) of components from different sources during sedimentary transport and sorting. For one sample, heavy-mineral fractionation may be indicated. These results have implications for studies of Nd model ages in sedimentary rocks, and for the interpretation of provenance from sedimentary petrography and geochemistry.

The neodymium model age of a sediment or sedimentary rock is generally interpreted as the average age of extraction from the mantle of the various provenance components¹. Complications involving uncertainties in mantle evolution models and possible changes in the Sm/Nd ratio during crustal processes² (such as partial melting) are potential sources of error in this interpretation. Recently, Nd model age data for sedimentary rocks have played an important role in modelling the evolution of the continental crust. For example, it has been noted that most terrigenous sedimentary rocks of post-Archaean age tend to have Nd model ages greatly in excess of their stratigraphic

ages^{1,3-6}. This has been interpreted in a number of ways. Veizer and Jansen^{7,8} emphasized the dominant role of cannibalistic sedimentary recycling in the formation of post-Archaean sediments, whereas others have emphasized the evolution of the contemporaneous upper continental crust^{3,4}. In spite of any differences in precise interpretation, most workers agree that Nd isotope composition provides an estimate of the mean provenance age.

One question that has received only scant attention is the possibility of a relationship between grain size and Nd isotope composition for sedimentary rocks. This is an important issue because in many cases the Nd isotope composition of individual samples (usually fine-grained) is generally taken as representative of the entire provenance. Goldstein *et al.*⁵ examined size fractions from a single sample of Amazon River sediment and found only trivial differences among the various size fractions. On the other hand, Frost and Winston⁹ examined coarse- and fine-grained sediments from the Proterozoic Belt-Purcell Supergroup of western North America and noted that the coarser-grained sediments had considerably more variable ϵ_{Nd} values and Sm/Nd ratios than did fine-grained sediment from the same sequence. They related this to changes in the Sm/Nd ratio during sedimentary recycling.

We are at present involved in a study of the geochemical and isotope composition of modern deep-sea turbidites from various tectonic settings¹⁰. Turbidites are an ideal sediment in which to examine any changes that might occur solely as a consequence of sedimentary sorting. For sediments that have not undergone diagenetic alteration, the change in grain size throughout most of a Bouma sequence is primarily related to a complex sequence of hydraulic sorting during transport and settling out of sediment from the water column. Sediment deposition from turbidity currents is comparatively rapid; however, depending on the time interval between turbidites, pelagic sediment may accumulate between cycles. As part of our study, a number of coexisting sands and muds from the same Bouma cycle were sampled.

The Nd isotope composition of a series of coexisting sands and muds from young turbidites from a variety of tectonic settings are listed in Table 1, along with relevant geochemical and petrographic data. The samples are the same as those used by Maynard and co-workers in their study of the petrographic and major-element composition of deep-sea turbidites¹¹⁻¹⁴. Although all major tectonic environments are represented, we concentrated on those that typically have abundant volcanogenic

Table 1 Neodymium isotope data

Sample	Sm* (p.p.m.)	Nd* (p.p.m.)	¹⁴⁷ Sm/ ¹⁴⁴ Nd	$f_{Sm/Nd}$	¹⁴³ Nd/ ¹⁴⁴ Nd	$\epsilon_{Nd}^{\dagger}$	$T_{DM}^{\ddagger}$	Tectonic setting	Area	Petrography (Q-F-L)§	SiO ₂ (wt.%)	CaO (wt.%)	Zr (p.p.m.)
V27-130-sand	2.25	10.9	0.125	-0.365	0.51130 (2)	-10.5	1.74	TE	Bay Biscay	87-9-4	84.2	5.01	228
V27-130-mud	7.25	36.2	0.121	-0.385	0.51133 (2)	-9.9	1.62				36.4	22.5	172
RC14-151-sand	4.41	18.2	0.146	-0.258	0.51223 (2)	+7.7	0.26	FA	S. Aleutians	0-20-80	55.3	7.07	116
RC14-151-mud	—	—	—	—	0.51205 (2)	+4.2	(0.64)¶				55.1	4.76	—
V28-327-sand	2.95	12.7	0.140	-0.288	0.51201 (4)	+3.4	0.67	BA	Celebes Basin	2-9-89	57.6	7.51	80
V28-327-mud	3.51	14.3	0.148	-0.248	0.51205 (2)	+4.2	0.66				55.8	7.21	96
RC12-374-sand	3.56	19.2	0.112	-0.431	0.51121 (3)	-12.2	1.65	BA	Japan	53-38-9	69.9	6.59	102
RC12-374-mud	6.12	32.7	0.113	-0.426	0.51136 (2)	-9.3	1.45				49.0	5.22	140
RC10-87-sand	3.46	17.9	0.117	-0.405	0.51155 (3)	-5.6	1.22	CA	Middle America	10-75-15	70.3	2.37	175
RC10-87-mud	5.55	26.0	0.129	-0.344	0.51190 (2)	+1.3	0.78				51.0	1.92	131
V28-357-sand	5.69	31.3	0.110	-0.441	0.51111 (2)	-14.2	1.76	CA	Sumatra	55-35-10	70.6	2.33	181
V28-357-mud	6.38	35.0	0.110	-0.441	0.51113 (2)	-13.8	1.73				61.1	1.88	157
V15-34-sand	4.95	28.3	0.106	-0.461	0.51167 (2)	-3.2	0.93	CA	Peru-Chile	33-49-18	68.0	4.40	192
V15-34-mud	4.89	23.7	0.125	-0.365	0.51158 (2)	-5.0	1.27				58.9	2.85	138

Errors in isotope composition are 2σ mean; Nd isotope composition normalized to $^{146}\text{Nd}/^{142}\text{Nd} = 0.636151$.

* Sm and Nd abundances by spark-source mass spectrometry with analytical uncertainties of ~5%; errors in Sm and Nd are correlated, resulting in uncertainties in Sm/Nd ratio of 2-4%.

† $\epsilon_{Nd} = [(^{143}\text{Nd}/^{144}\text{Nd})_s / (^{143}\text{Nd}/^{144}\text{Nd})_n - 1] \times 10^4$ and $f_{Sm/Nd} = [(^{147}\text{Sm}/^{144}\text{Nd})_s / (^{147}\text{Sm}/^{144}\text{Nd})_n - 1]$, where $(^{143}\text{Nd}/^{144}\text{Nd})_n = 0.511836$ and $(^{147}\text{Sm}/^{144}\text{Nd})_n = 0.1967$; s and n represent sample and normalizing (chondritic reservoir; CHUR) values respectively.

‡ $T_{DM} = 1/\lambda_{Sm} \{1 + [(^{143}\text{Nd}/^{144}\text{Nd})_s - (^{143}\text{Nd}/^{144}\text{Nd})_{DM}] / [(^{147}\text{Sm}/^{144}\text{Nd})_s - (^{147}\text{Sm}/^{144}\text{Nd})_{DM}]\}$, where $(^{143}\text{Nd}/^{144}\text{Nd})_{DM} = 0.51235$ and $(^{147}\text{Sm}/^{144}\text{Nd})_{DM} = 0.217$ and $\lambda_{Sm} = 6.54 \times 10^{-12} \text{ yr}^{-1}$. s and DM represent sample and depleted-mantle values respectively.

§ Per cent quartz, feldspar and unstable lithic fragments.

|| TE: trailing edge; FA: fore-arc; BA: back-arc; CA: continental arc.

¶ Model age of RC14-151-mud calculated using same $^{147}\text{Sm}/^{144}\text{Nd}$ ratio as RC14-151-sand.

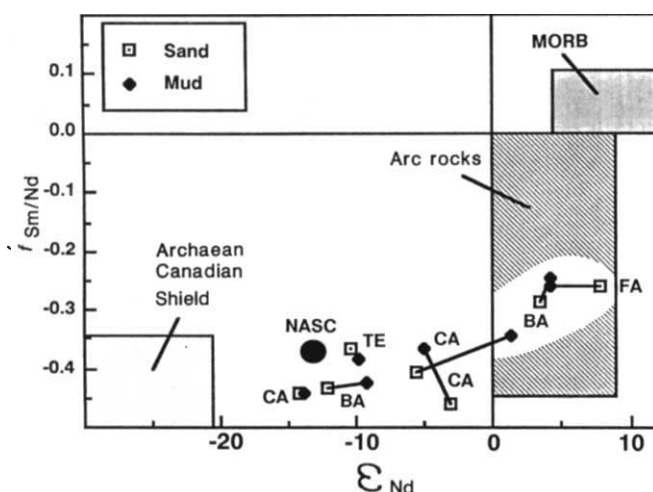


Fig. 1 Plot of $f_{\text{Sm/Nd}}$ against ϵ_{Nd} for sand-mud pairs. Two-component mixing is represented by straight lines¹⁵. Shown for comparison are generalized fields for mid-ocean-ridge basalts (MORBs) and typical island-arc and continental-arc igneous rocks, compiled from the literature. Also shown are the field for composite samples from the Archaean-aged Canadian Shield¹ and the values for the North American shale composite (NASC)¹. It is suggested that much of the difference seen in some sand-mud pairs can be related to a separation of various provenance components. Such separation may be accomplished by original grain-size contrasts of provenance components, grain-size differences generated by weathering processes preferentially affecting certain lithologies and, in some cases, by heavy mineral fractionation during sedimentary transport. CA denotes continental-arc rocks; for island-arc rocks, TE denotes trailing edge, FA denotes fore-arc and BA denotes back-arc.

material in the provenance, because these were considered the most likely candidates to exhibit significant differences in Nd isotopes (see below). In all samples, the coarse- and fine-grained portions come from the same turbidite horizon.

There are a number of features worth noting. First, the muds have significantly higher Nd abundances (from 13% to over 300%) than do the sands in all but one pair. This is a not uncommon feature of rare-earth-element (REE) abundances in sediments² and generally can be attributed to a dilution effect from quartz. Second, whereas Sm/Nd ratios of coexisting sands and muds agree to within 20%, there are significant differences for two of the sample pairs (Sm/Nd ratios are accurate to only ~2–4%). Unlike the differences noted by Frost and Winston⁹, where coarse sediment had higher Sm/Nd ratios, in this case the coarse fractions generally have lower Sm/Nd ratios.

By far the most important feature is that in four of seven sample pairs, the sands and muds have clearly different Nd isotope compositions of more than 1 ϵ_{Nd} unit. This results in different Nd model ages and, accordingly, would indicate distinct provenance and recycling histories^{3–9}, regardless of the exact cause. One sample from the Middle Americas has a difference of nearly 7 ϵ_{Nd} units. Of the four pairs showing a difference in Nd isotope composition of greater than 1.0 ϵ_{Nd} units, the muds have higher $^{143}\text{Nd}/^{144}\text{Nd}$ ratios (that is, more positive ϵ_{Nd}) in two of the pairs and lower $^{143}\text{Nd}/^{144}\text{Nd}$ ratios in the other two. For two additional sample pairs with differences of 0.6 and 0.8 ϵ_{Nd} units (which are not necessarily significant), the muds have higher $^{143}\text{Nd}/^{144}\text{Nd}$ ratios. The two samples that show significant differences in Sm/Nd ratios for coexisting sand and mud also show substantial differences in Nd isotope composition.

Considerably more data than that collected for this study will be required to fully understand the complex Nd isotope variations reported here. It seems likely, however, that the very

different Nd isotope composition (and in two cases different Sm/Nd ratio) observed for sands and muds suggests that a separation (or 'unmixing') of provenance components may have taken place during sedimentary transport. In Fig. 1, the data are displayed on a plot of $f_{\text{Sm/Nd}}$ against ϵ_{Nd} . Shown for comparison are fields encompassing most mid-ocean-ridge basalts (MORBs), typical island-arc igneous rocks, results for composite samples from the Archaean portion of the Canadian Shield¹ and the North American shale composite (NASC)¹. Although the provenance history of many of these sediments is undoubtedly very complex, it can be assumed for simplicity that the samples considered here are derived from variable mixtures of old recycled upper crust (the Archaean Shield would be an extreme example and NASC is perhaps a more typical example) and young, arc-derived volcanics. On such a diagram, two-component mixing is represented by a straight line¹⁵.

There are a number of mechanisms that could result in separation of provenance components according to sediment grain size. Original contrasts in grain size of provenance components could result in finer-grained material (such as volcanics) being incorporated preferentially in the mud fraction. A second mechanism is the preferential breakdown of certain components during weathering and transport^{16,17}. In principle, this should tend to affect most strongly finer-grained and/or labile source rocks (that is, those containing glass and unstable mineralogy), resulting in a concentration of such material in finer-grained sediment. A final consideration is the preferential enrichment of heavy minerals in coarser-grained sediment during sedimentary transport and sorting, a result of simple hydraulic-equivalence considerations. A number of minor minerals, such as zircon and monazite^{2,18} could significantly affect the REE (and consequently the Nd isotope) mass balance.

For an unmixing model in which volcanogenic material is concentrated in the muds, the finer-grained material should have higher ϵ_{Nd} values. In Fig. 1, this is consistent with five of the sample pairs, although the difference is only statistically distinct for two. The remaining two sand-mud pairs exhibit the opposite trend (RC14-151, Aleutians, fore-arc; V15-34, Peru-Chile, continental arc). RC14-151 is composed virtually entirely of volcanogenic material, with framework grains in the sand consisting of no free quartz and about 80% volcanic lithic grains. Accordingly, the discrepancy in Nd isotope composition does not easily fit into the simple two-component model described above and may be related to 'unmixing' or other processes taking place between various components within the volcanic provenance. Sample V15-34 is clearly derived from a mixed provenance, with the sand having a high content of both quartz (33%) and volcanic, lithic rock fragments (18%). This sample is distinct from the others for a number of reasons. It is one of only two samples that has a significant difference in Sm/Nd ratio for sand and mud (and is the sample with the largest difference) and is the only sample for which the sand has a higher Nd concentration (as well as other light REEs, Th and U) than the mud. The Zr content is only slightly different (Table 1). These observations may indicate that heavy minerals (presumably other than or in addition to zircon) are playing an important role in the differing Nd isotope composition of this sample pair. There is no clear indication of any specific mineral from other geochemical data, possibly indicating a light-REE-rich phase, such as monazite.

Of the four sample pairs with the largest differences in Nd isotope composition, three have a substantial volcanic component, evident in the sandstone petrography. In the other sample (RC12-374, Japan, back-arc), sandstone petrography reveals a smaller volcanic component, but other geochemical data suggest that the mud may have a considerably greater volcanic component. Only one of the sample pairs with a major volcanic component does not show a significant difference in Nd isotope composition (V28-327, Celebes Basin, back-arc). This sample is also distinct in that the overall major- and

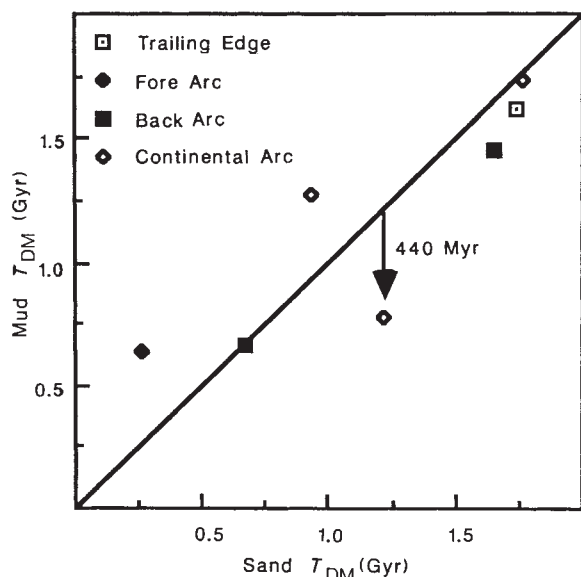


Fig. 2 Plot of neodymium model ages (relative to depleted mantle) for coexisting sands and muds. The diagonal line represents equivalent ages. Age differences of up to 440 Myr are observed for some sand-mud pairs. Differences tend to be greatest for samples with lower model ages. Some care must be exercised in interpreting the fore-arc sample pair (RC14-151), because the Sm/Nd ratio of the mud was assumed to be the same as that for the sand (see Table 1).

trace-element chemistry of the sand and mud is virtually identical (note, for example, the SiO₂ data in Table 1), with only slight differences in the amount of carbonate. The sand fraction is also particularly fine-grained and it is likely that relatively minor grain-size fractionation has occurred in this sample pair.

The implications for Nd model ages can be seen in Fig. 2, in which the model ages of sand and mud are plotted against each other. It appears that the sample pairs with the lowest model ages are most likely to show the greatest differences between sand and mud. We consider this to be consistent with preferential breakdown of a young volcanic component, as described above. The difference in model age between sand and mud is as much as 0.44 Gyr (with mud being either higher or lower), and would clearly be significant for many model age studies. On the other hand, we do not necessarily regard this phenomenon as seriously affecting all sedimentary sequences. We suggest that the weathering and transport of a highly labile component, typically from a first-cycle volcanic provenance, are the most efficient methods for causing significant differences in the Nd isotope composition, by an unmixing mechanism, during sedimentary sorting. Accordingly, we would expect that sediments that have seen a long recycling history (that is, the majority of preserved sediment^{7,8}), would be less likely to show substantial differences in Nd isotope composition during sorting.

These results have a bearing on provenance studies based on sandstone petrography. Recently, there has been a great deal of effort to relate the petrography of sandstones to the provenance and tectonic setting of deposition (see, for example, refs 19, 20). The data reported here indicate a significant separation of provenance among various grain sizes, a conclusion supported by much petrographic data for sandstones (ref. 21, for example). The Nd isotope data suggest that this phenomenon may affect volcanic sources preferentially over plutonic or recycled sources. Thus, failure to take into account all grain sizes may lead to an incomplete understanding of provenance.

We are grateful to Sidney Hemming and Eirik Krogstad for commenting on the manuscript and to Steve Goldstein for providing especially helpful reviews.

Received 15 July; accepted 16 December 1988.

- McCulloch, M. T. & Wasserburg, G. J. *Science* **200**, 1003-1011 (1978).
- Taylor, S. R. & McLennan, S. M. *The Continental Crust: Its Composition and Evolution* (Blackwell, Oxford, 1985).
- O'Nions, R. K., Hamilton, P. J. & Hooker, P. J. *Earth planet. Sci. Lett.* **63**, 229-240 (1983).
- Allègre, C. J. & Rousseau, D. *Earth planet. Sci. Lett.* **67**, 19-34 (1984).
- Goldstein, S. L., O'Nions, R. K. & Hamilton, P. J. *Earth planet. Sci. Lett.* **70**, 221-236 (1984).
- Taylor, S. R., McLennan, S. M. & McCulloch, M. T. *Geochim. cosmochim. Acta* **47**, 1897-1905 (1983).
- Veizer, J. & Jansen, S. L. *J. Geol.* **87**, 341-370 (1979).
- Veizer, J. & Jansen, S. L. *J. Geol.* **93**, 625-643 (1985).
- Frost, C. D. & Winston, D. *J. Geol.* **95**, 309-327 (1987).
- McLennan, S. M., McCulloch, M. T., Taylor, S. R. & Maynard, J. B. *Eos* **66**, 1136 (1985).
- Valloni, R. & Maynard, J. B. *Sedimentology* **28**, 75-83 (1981).
- Maynard, J. B., Valloni, R. & Yu, H.-S. in *Trench and Fore-arc Sedimentation* (ed. Leggett, J. K.) 551-561 (Geological Society, London, 1982).
- Maynard, J. B. *Sedimentology* **31**, 493-501 (1984).
- Yerino, L. N. & Maynard, J. B. *Sedimentology* **31**, 83-89 (1984).
- Shirey, S. B. & Hanson, G. N. *Geochim. cosmochim. Acta* **50**, 2631-2651 (1986).
- Garrels, R. M. & Mackenzie, F. T. *Evolution of Sedimentary Rocks* (Norton, New York, 1971).
- Nesbitt, H. W. & Young, G. M. *Geochim. cosmochim. Acta* **48**, 1523-1534 (1984).
- Cullers, R., Chaudhuri, S., Kilbane, N. & Koch, R. *Geochim. cosmochim. Acta* **43**, 1285-1301 (1979).
- Dickinson, W. R. & Suczek, C. A. *Bull. Am. Ass. Petrol. Geol.* **63**, 2164-2182 (1979).
- Potter, P. E. *Nature* **311**, 645-648 (1984).
- Potter, P. E. *J. Geol.* **86**, 423-449 (1978).

A parallel algorithm for real-time computation of optical flow

H. Bülthoff*, J. Little & T. Poggio

Center for Biological Information Processing at the Brain and Cognitive Science Department, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The precise measurement of the two-dimensional field of velocities from time-varying two-dimensional images is impossible in general. It is, however, possible to compute suitable 'optical flows' that are qualitatively similar to the velocity field in most cases. We describe a simple, parallel algorithm that computes an optical flow from sequences of real images, which is consistent with human psychophysics and suggests plausible physiological models. In particular, our algorithm runs on a Connection Machine supercomputer in close-to-real time. It shows several of the same 'illusions' that are perceived by humans. A natural physiological implementation of the model is consistent with data from cortical areas V1 and MT.

If an observer moves in a three-dimensional world, a flow of apparent motion (optical flow) occurs on his retina which is useful for performing several different tasks such as image segmentation, determination of structure from motion, and navigation. We have developed a new parallel algorithm to compute the optical flow in near-real time for natural images (256 × 256 pixels) on the Connection Machine system. The algorithm relies on the assumption that the optical flow is due locally to a first approximation to frontoparallel translation of a lambertian (matte) surface. It can be described as follows (Fig. 1): consider a set of layers of processors holding the results of multiplying or logical 'anding' image 'features' (either edges or image intensity values, or filtered intensity values) for different displacements of one image relative to the next, each layer corresponding to a different displacement. The displacements can be different in magnitude and direction and each displacement corresponds to a discrete velocity $v(x, y)$. Each processor sums excitatory inputs from processors in a small excitatory neighbourhood (typically 11 by 11 pixels) in the same layer: a large total excitation is expected if the motion field is locally uniform at that velocity. Next, for each (x, y) position, the processor that has the maximum excitation among all processors in the other layers is selected. This is the non-maximum suppression or winner-take-all step. The corresponding $v(x, y)$ is taken as

* Present address: Department of Cognitive and Linguistic Sciences, Brown University, Providence, Rhode Island 02912, USA.

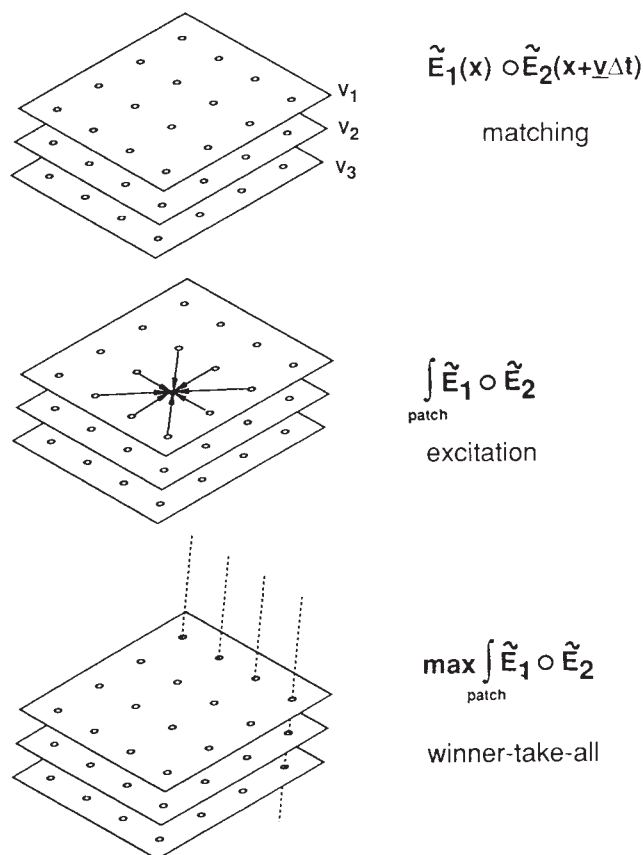


Fig. 1 A schematic representation of the three main stages of the algorithm: the network of processors is depicted on the left, with the corresponding mathematical operations on the right. The first step is matching of features or intensity values from two successive frames: the matching operation could be multiplication between intensity values (or filtered intensities), as well as a logical operation between binary features such as edges. This is done for several possible displacements between the two images, corresponding to different velocities. The second step is the local summation step (excitation): each processor in each layer gets support from a fixed neighbourhood in the same layer (for the simple case of binary features; the general case is similar^{18,21}). The third step finds the velocity corresponding to the maximum excitation for each position x, y . The network representation on the left is a quite faithful description of the Connection Machine implementation. The formulae on the right show that the algorithm is similar to patchwise correlation (with overlap) which can be regarded as a form of regularization²¹.

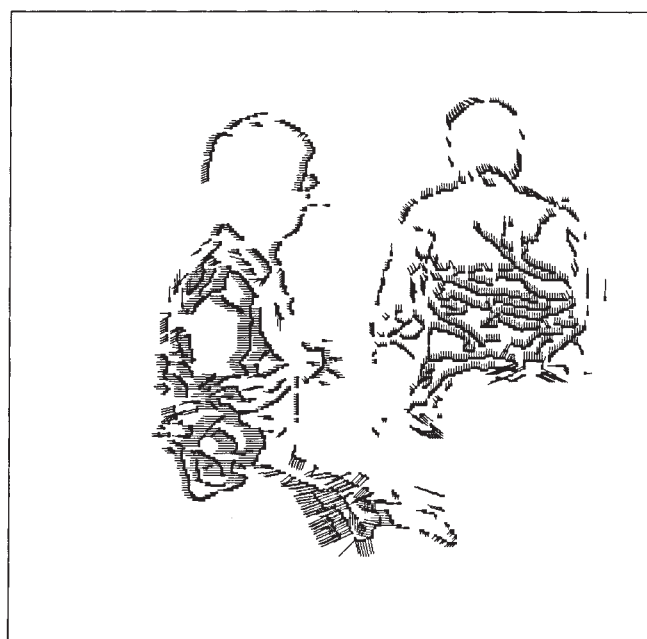


Fig. 2 A natural sequence of two frames (only one shown here) and the optical flow computed by the algorithm. The person on the right is standing up (translation) and the person on the left is turning away from the Connection Machine monitor (rotation). The range of possible motions was restricted to ± 6 pixels in any direction and the excitatory region was the same size. In this example the features used for matching were intensity edges²².

the velocity of the point (x, y) . This scheme is similar to a Connection Machine implementation of a stereo algorithm¹.

The algorithm has been implemented as part of the Vision Machine system². It has been intensively tested with synthetic and natural images using one of several different types of features or measurements on the intensity data, including edges (both zero-crossings of the laplacian of gaussian and Canny's method) (Fig. 2), or intensity data appropriately filtered (for instance, through a centre-surround operator).

The algorithm is consistent with several perceptual effects³. For instance, it suffers from the so-called 'barber-pole illusion' in the same way as humans do. Figure 3a shows that whereas the true velocity vectors for a rotating barber's pole are strictly horizontal, our algorithm computes a vertical velocity field consistent with the illusion. This is also a plausible physical interpretation for the optical flow field in the image. Therefore it is not surprising that other motion algorithms are also consistent with perception⁴. Note, however, that we do not require global smoothing (in this case the diameter of the excitatory region is 13 pixels and the diameter of the barber's pole is 128 pixels).

We could also simulate a psychophysical observation described recently by Nakayama and Silverman⁵. They addressed the question of whether, to solve the aperture problem, early motion measurements are integrated over areas (as suggested by Horn and Schunck⁶ and by our algorithm) or integrated along contours (as proposed by Hildreth⁴). Their study uses a simple distorted line moving up and down on an otherwise untextured background (Fig. 3b). When the distortion is small and the line is almost straight, the central diagonal section seems to move in an oblique direction, whereas the straight parts of the line move up and down. This leads to a non-rigid perception of the entire figure. The figure can be made to move more rigidly both perceptually and in the algorithm by

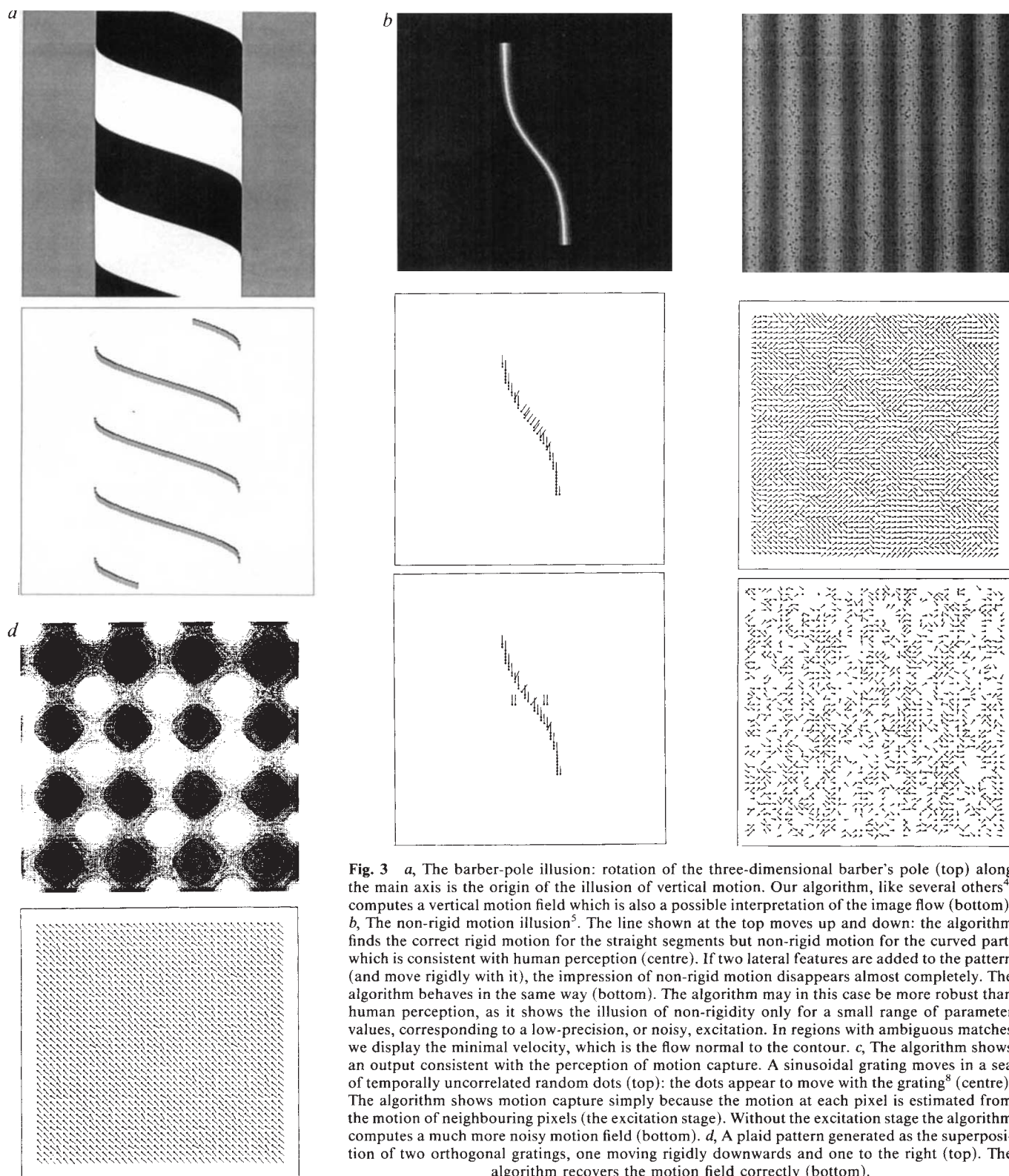


Fig. 3 *a*, The barber-pole illusion: rotation of the three-dimensional barber's pole (top) along the main axis is the origin of the illusion of vertical motion. Our algorithm, like several others⁴, computes a vertical motion field which is also a possible interpretation of the image flow (bottom). *b*, The non-rigid motion illusion⁵. The line shown at the top moves up and down: the algorithm finds the correct rigid motion for the straight segments but non-rigid motion for the curved part, which is consistent with human perception (centre). If two lateral features are added to the pattern (and move rigidly with it), the impression of non-rigid motion disappears almost completely. The algorithm behaves in the same way (bottom). The algorithm may in this case be more robust than human perception, as it shows the illusion of non-rigidity only for a small range of parameter values, corresponding to a low-precision, or noisy, excitation. In regions with ambiguous matches we display the minimal velocity, which is the flow normal to the contour. *c*, The algorithm shows an output consistent with the perception of motion capture. A sinusoidal grating moves in a sea of temporally uncorrelated random dots (top): the dots appear to move with the grating⁸ (centre). The algorithm shows motion capture simply because the motion at each pixel is estimated from the motion of neighbouring pixels (the excitation stage). Without the excitation stage the algorithm computes a much more noisy motion field (bottom). *d*, A plaid pattern generated as the superposition of two orthogonal gratings, one moving rigidly downwards and one to the right (top). The algorithm recovers the motion field correctly (bottom).

increasing the curvature of the sigmoid part of the line, or by introducing additional features on or off the line that are unambiguously moving up and down (Fig. 3*b*). Our algorithm agrees with the psychophysical observations qualitatively. Of course the rigidity depends on the excitatory region that has to be larger for less curved lines⁷. Figure 3*c* shows the perceptual illusion termed 'motion capture'⁸. If a moving sine-wave grating is superimposed on a pair of alternating and uncorrelated random-dot patterns, most of the dots in the display appear to move as a uniform sheet in synchrony with the sine-wave grating. The

algorithm shows the same effect, in this simple case, without using any of the tricks suggested by Ramachandran⁹. Motion of a 'plaid' pattern of two superimposed sinusoidal gratings is also correctly recovered by the algorithm (Fig. 3*d*), which is consistent with perception.

Our algorithm is also consistent with a modification of Wallach's aperture experiment¹⁰. The algorithm finds a vertical motion field if an obliquely oriented grating moves horizontally behind a vertical slit (Fig. 4*a*). It finds horizontal motion when the slit is horizontal (Fig. 4*b*). The excitatory region has to cover

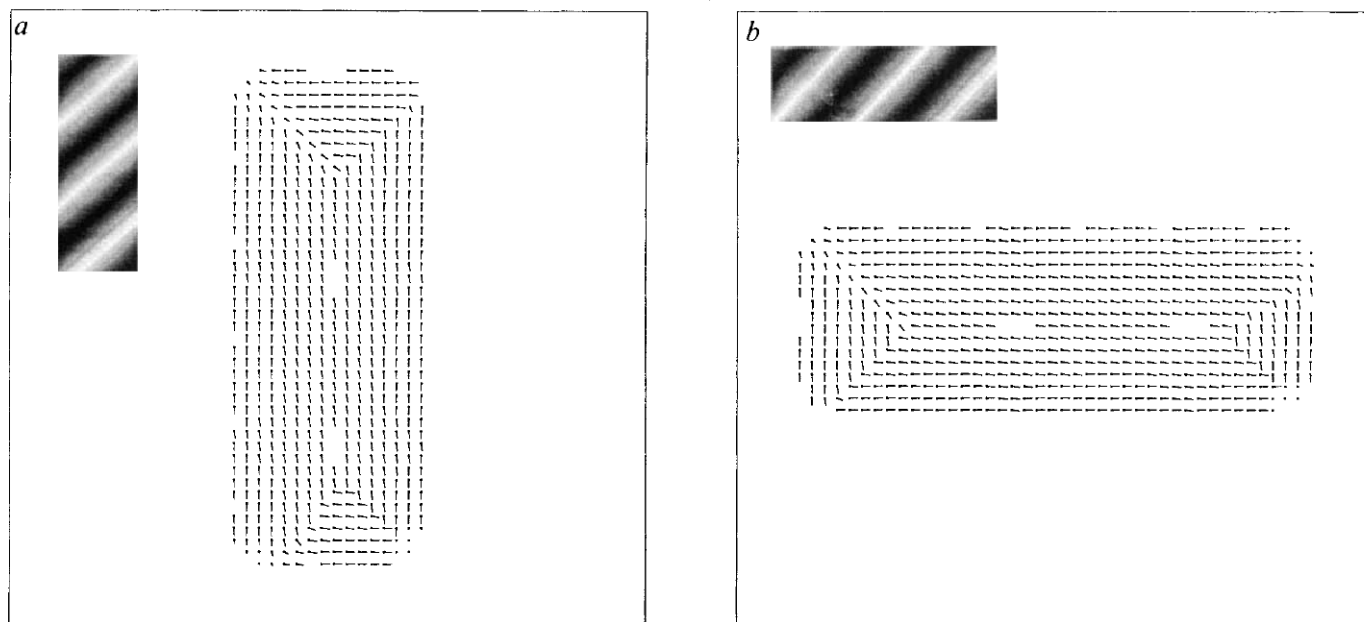


Fig. 4 *a*, An oblique grating moves horizontally behind a vertical window. The algorithm computes an almost vertical flow field, which is consistent with perception. If the window is horizontal, the same motion leads to a horizontal flow field (*b*), again consistent with perception.

the whole width of the aperture (otherwise only the regions close to the slit boundary can be disambiguated). An iterative version of our algorithm may propagate the information from the boundaries of the window to the inside of the grating, progressively disambiguating the velocity field, even with a small excitatory region.

It is natural to ask whether some version of this algorithm may have a plausible implementation in terms of cortical physiology. There are psychophysical data suggesting that motion is computed in two stages, with the first stage computing the perpendicular components and the second stage combining these measurements over an extended area into a coherent motion pattern¹¹. There is also physiological evidence for a two-stage motion computation in primate visual cortex. Movshon *et al.*¹² found that motion-sensitive neurons in area V1 could only compute the component of motion in the direction perpendicular to the orientation of image features. These neurons responded to only one component of two superimposed sine-wave gratings moving in different directions (same stimulus as used in ref. 11). In area MT, however, cells have been found that seem to respond to the direction of motion of the combined pattern (pattern cells). These psychophysical and physiological observations suggest a physiological implementation of the parallel network of Fig. 1. The first step of the algorithm, the matching stage, may be implemented by cortical motion detectors such as the direction-selective cortical cells in area V1, or the elementary motion detectors in the fly visual system¹³. The underlying circuitry is an open question, although it could be similar to the one proposed for disparity-sensitive cells and based on silent inhibition¹⁴. The summation stage may be carried out by a set of neurons with a large receptive field that collects information from direction selective cells in V1 and may correspond to the voting neighbourhood of Fig. 1. These neurons are tentatively identified with neurons in MT. The last step, the winner-take-all step, could be implemented by inhibitory connections between neurons in MT tuned to different velocities but with the same receptive field location. Our algorithm has much simpler and more natural physiological implementations than other motion schemes discussed recently^{15,16}.

The algorithm described here suffers from the aperture problem in only a mild form. More precisely, it will find a unique optical flow field except in pathological cases (a very long straight contour with no visible terminations or features). The

comparison stage is similar to patchwise cross-correlation of features (for a related method see ref. 17). It treats intensity values and edges in a uniform way. For matching features, we have used zero-crossings, the laplacian of gaussian filtered images, their sign and the smoothed intensity values, with similar results. It is not surprising that methods which are superficially so different (edge-based and intensity-based) give such similar results, because there are well known theoretical arguments supporting the equivalence of cross-correlating the sign bit of the laplacian filtered image and the laplacian filtered image itself¹⁸.

The algorithm suggests an interesting general point: phenomena such as motion capture are to be expected by any algorithm that integrates information about motion over local spatial neighbourhoods, possibly with spatially varying weights, as do standard regularization algorithms such as those of Horn and Schunck⁶ and a recent scheme of Yuille and Grzywacz¹⁶. It also suggests that the size of the neighbourhood is critically important and should depend on the image data. The question is still open as to how the best neighbourhood size should be chosen and in particular how the human visual system does it.

The algorithm by itself is clearly not a theory of motion perception, which is likely to require more than one single module. In the Vision Machine system, the motion algorithm is used in conjunction with later stages which find motion discontinuities, interpolate and smooth the optical flow and integrate it with other visual modalities¹⁹. Finally, we are currently working towards a VLSI implementation of the algorithm, using charge-coupled-device technology²⁰.

This research is supported by the Office of Naval Research, Engineering Psychology Division, the National Science Foundation, the Artificial Intelligence Center of Hughes Aircraft Corporation, the NATO Scientific Affairs Division and the Department of Defence. T.P. is supported by the Uncas and Ellen Whitaker chair. Patrick A. O'Donnell helped us with the computer simulations on the Connection Machine.

Received 3 August; accepted 20 December 1988.

1. Drumheller, M. & Poggio, T. in *Proc. IEEE Conf. on Robotics and Automation* 1439-1448 (1986).
2. Poggio, T. *et al.* in *Proc. Image Understanding Workshop* (Morgan Kaufmann, San Mateo, California, 1988).
3. Bülthoff, H. H., Little, J. J. & Poggio, T. *J. opt. Soc. Am.* **4**, 34 (1987).
4. Hildreth, E. C. *The Measurement of Visual Motion* (MIT, Cambridge, Massachusetts, 1984).

5. Nakayama, K. & Silverman, G. H. *Vision Res.* **28**, 739-746 (1988).
6. Horn, B. K. P. & Schunck, B. G. *Artif. Intell.* **17**, 185-203 (1981).
7. Bülthoff, H. H. & Little, J. J. in *Motion Workshop* (IEEE, Washington, DC, in the press).
8. Ramachandran, V. & Inada, Y. *Spatial Vision* **1**, 57-67 (1985).
9. Ramachandran, V. S. & Anstis, S. M. *Scient. Am.* **254**, 102-109 (1986).
10. Wallach, H. in *On Perception* (Quadrangle, New-York, 1976).
11. Adelson, E. & Movshon, J. *Nature* **300**, 523-525 (1982).
12. Movshon, J. A., Adelson, E. H., Gizzi, M. S. & Newsome, W. T. in *Pattern Recognition Mechanisms* (ed. Chagas, C. G. G. & Gattas, R.) (Vatican, Rome, 1985).
13. Hassenstein, B. & Reichardt, W. *Z. Naturf.* **11b**, 513-524 (1956).
14. Poggio, G. F. & Poggio, T. A. *Rev. Neurosci.* **7**, 379-412 (1984).
15. Hutchinson, J., Koch, C., Luo, J. & Mead, C. *IEEE Computer Magazine* **21**, 52-64 (1988).
16. Yuille, A. L. & Grzywacz, N. M. *Nature* **333**, 71-74 (1988).
17. Nishihara, H. *Optical Engng* **23**, 536-545 (1984).
18. Little, J. J., Bülthoff, H. H. & Poggio, T. in *Proc. Image Understanding Workshop* 915-920 (Morgan Kaufmann, San Mateo, California, 1987).
19. Poggio, T., Gamble, E. B. & Little, J. J. *Science* **242**, 436-440 (1988).
20. Horn, B. et al. *Technical report* (Artificial Intelligence Laboratory, MIT, 1988).
21. Little, J. J., Bülthoff, H. H. & Poggio, T. in *Proc. Int. Conf. Computer Vision* (IEEE, Washington, DC, 1988).
22. Canny, J. F. *IEEE Trans. Pattern Analysis and Machine Intelligence* **8**, 679-698 (1986).

Intrinsic differences in the growth rate of early nerve fibres related to target distance

Alun M. Davies

Department of Anatomy, St George's Hospital Medical School, Cranmer Terrace, Tooting, London SW17 0RE, UK

Target field innervation in the developing vertebrate nervous system coincides with the onset of important trophic interactions¹. Two factors that determine the timing of this event are the distance axons have to grow to reach their targets, which are known to vary, and the rate at which they grow. There have been few studies of axonal growth rate at this stage of development²⁻⁴ and no comparative study of the relationship between growth rate and target distance. Embryonic chick cranial sensory neurons are located in discrete ganglia and the distance axons have to grow to reach their targets is different for each ganglion, ranging from several hundred to several thousand μm . Here, I show that these neurons differ in their *in vivo* growth rates; neurons with more distant targets growing faster. *In vitro*, single isolated neurons from each of these populations grow at a similar rate to that observed *in vivo*, indicating that growth rate is an intrinsically determined property of neurons before they reach their targets.

To compare the growth rate of different neurons at the same stage of their development, I used the vestibular, geniculate, petrosal and nodose ganglia whose neurons are born over the same period of development (between two and five days *in ovo*⁵). These ganglia also have similar ontogeny, in that their neurons are derived exclusively from ectodermal placodes⁶. The distance between the central nervous system and the peripheral targets of these ganglia increases from vestibular through geniculate and petrosal to nodose (Fig. 1).

The rate at which the neurons of each ganglion grow *in vivo* was estimated from measurements of nerve lengths in a staged series of silver-stained embryo whole-mounts. Vestibular neurons have the slowest growth rate, geniculate and petrosal neurons are between two and threefold faster, and nodose neurons are fivefold faster (Fig. 2). The order is the same for the distances that the neurons of these ganglia have to grow to reach their targets in the embryo: the further the targets, the faster the neurons grow.

To determine whether differences in growth rate are intrinsic properties of the neurons or are due to differences in the environment through which their axons grow, the growth of sensory neurons from each ganglion was studied *in vitro* using an identical laminin substratum in each set of cultures. The ganglia were dissected from embryos at stages during the period in which neuroblasts undergo their terminal division, when their peripheral and central axons are growing to their targets in the periphery and CNS. Dissociated, low-density cell cultures were set up and monitored for 18 hours. In these cultures, neurons had a bipolar morphology with fine, unbranched neurites growing in opposite directions from small, spindle-shaped cell bodies (Fig. 3). The neurons of each ganglion grew at a distinct rate similar to that observed *in vivo* (Fig. 4a). These differences in growth rate cannot be ascribed to differences in the developmental stage of the neurons because they are born during the same period of development⁵ and extend axons to their targets at the same time (Fig. 2). Furthermore, similar results were obtained in cultures set up at stages during the period in which neurons are born in these ganglia, namely, at stages 20, 22 and 24 (data not shown). Thus, these *in vitro* results indicate that growth rate is an intrinsically regulated characteristic of early neurons. To exclude the possibility that neuronal growth rate was influenced by substances released into the culture medium by other cells, single-cell cultures were set up in multiwell plates. Similar results were observed in these cultures (Fig. 4b).

Later in development, when neurons innervate their targets, their survival becomes dependent on target-derived neurotrophic factors^{7,8}. Brain-derived neurotrophic factor promotes the survival and growth of vestibular, geniculate, petrosal and nodose neurons at this later stage⁹, and has no effect on the growth rate of these neurons before they reach their targets (data not shown).

In previous studies, rates of between 20 and 75 $\mu\text{m h}^{-1}$ for the growth of axons to their targets *in vivo* have been suggested²⁻⁴. Because data were available for only a handful of neuronal populations, relationships between growth rate and target distance were not unrecognized. Although there has been much more work on the growth of several different kinds of embryonic neurons *in vitro*¹⁰⁻¹³, there has been no systematic comparative study of the growth rates of different populations of neurons under similar conditions. Furthermore, almost all of this *in vitro* work used neurons that had already innervated their

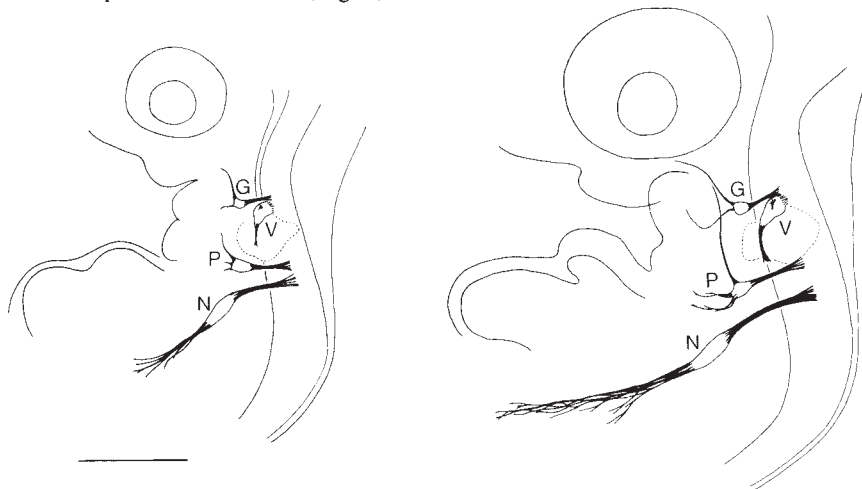


Fig. 1 Camera lucida drawings of wholemount silver-stained chick embryos¹⁴ at stages 22 (left) and 24 (right) (about 3.5 and 4 days of development) to show the location of the cranial sensory ganglia used in this study together with the course of their central and major peripheral nerve branches. The central axons are drawn to the point where they enter the brainstem, the relatively short course of these fibres within the brainstem cannot be clearly resolved in silver stained preparations. For clarity, the somatic efferent and branchial efferent fibres which leave the brainstem and run with the sensory nerves for a short part of their course are not shown. G, geniculate ganglion; V, vestibular ganglion; P, petrosal ganglion; N, nodose ganglion. Scale bar, 1 mm.

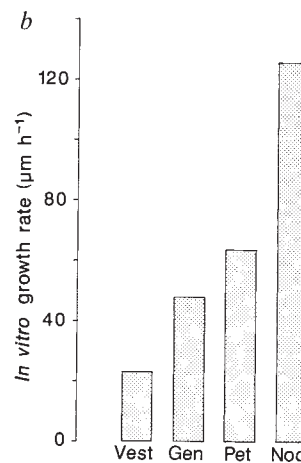
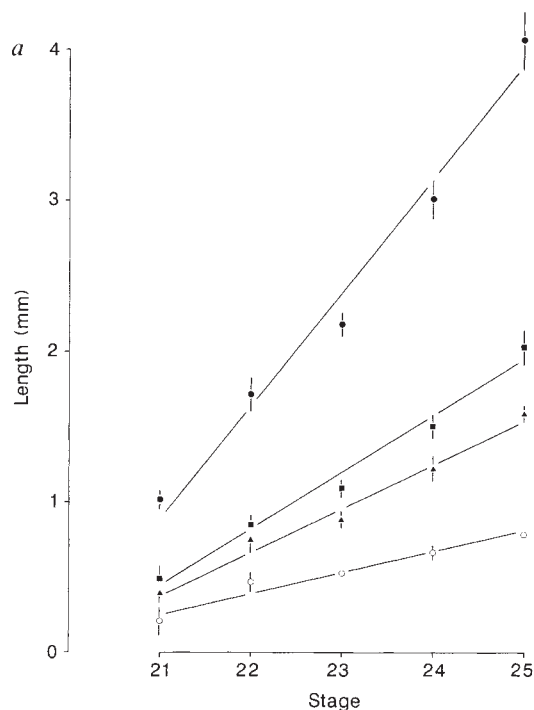


Fig. 2 Estimates of the lengths and growth rates of cranial sensory neurons *in vivo*. The distances from the most distal growth cones in the longest peripheral nerve branches of each ganglion to the point of entry of the central axons into the brainstem were measured from *camera lucida* drawings of silver-stained whole-mount embryos. *a*, Distances from stages 21 to 25 (about six-hourly intervals). ○, vestibular; ▲, geniculate; ■, petrosal; ●, nodose. The mean $\pm$ s.e.m. of measurements taken from at least three embryos at each stage are shown together with the lines of best fit. *b*, Bar chart of growth rates derived from these data. Bar height refers to the rate of elongation of neurons (the combined growth rate of peripheral and central axons).

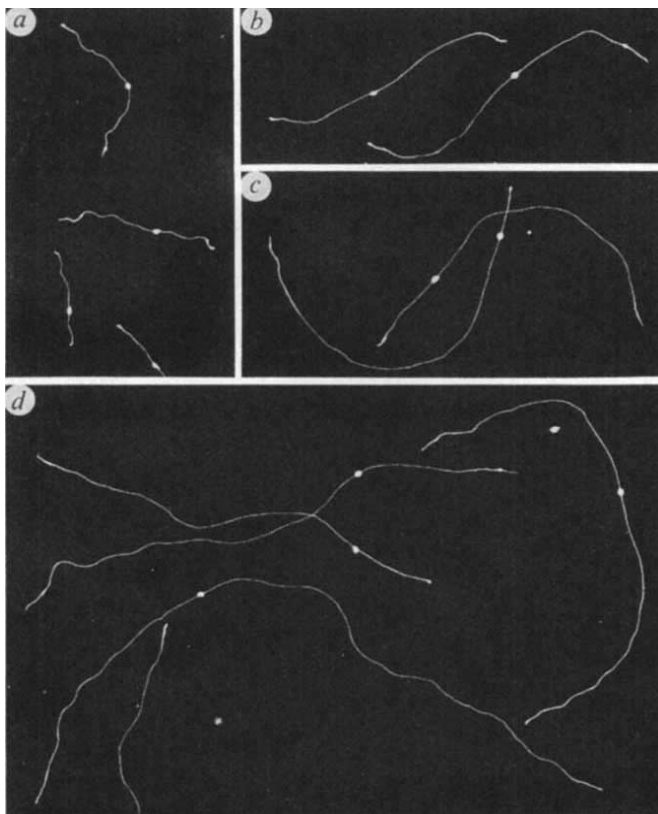


Fig. 3 Photomontages of vestibular (*a*), geniculate (*b*), petrosal (*c*) and nodose (*d*) ganglion neurons after 12 hours in culture. Scale bar, 100 μ m.

Methods. Stage 22 ganglia were trypsinized, dissociated by trituration and plated on a poly ornithine/laminin substratum in Nunc plastic tissue culture dishes and grown at 37 °C in F14 medium supplemented with 10% horse serum and 5% FCS. To show the extent of the fine neurites of these early neurons more clearly, standard immunocytochemical techniques were used to stain the cultures with a rabbit anti-68K neurofilament antibody (gift of P. Hollenbeck) followed by an fluorescein isothiocyanate-conjugated goat anti-rabbit IgG secondary antibody.

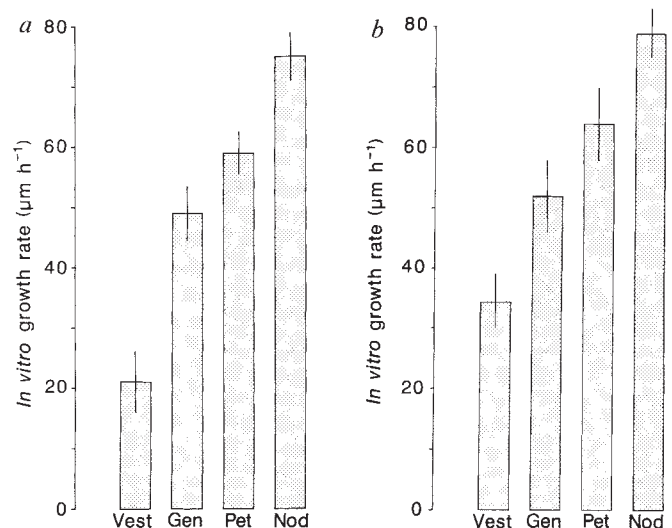


Fig. 4 *a*, Bar chart of the growth rate of stage 22 vestibular, geniculate, petrosal and nodose ganglion neurons cultured for 18 hours; the mean $\pm$ s.e.m. of the growth rates of 50 neurons in each culture is shown. *b*, Bar chart of the growth rate of stage 22 neurons cultured for 18 hours as isolated cells in 16-mm wells (Nunc multiwell plates); the mean $\pm$ s.e.m. of the growth rate of 20 neurons of each ganglion is shown. Note that bar height in *a* and *b* refers to the rate of elongation of neurons (the combined growth rate of both neurites that extend from the cell body).

Methods. The growth rate of each neuron was calculated from serial, two-hourly measurements of its total length. Measurements of a random sample of neurons at intervals would not provide an accurate estimate of growth rate because neurons continue to differentiate from progenitors in these cultures throughout the period of study. The neurons were grown in 60 mm diameter culture dishes at a density of about 500 cells per dish. For serial measurements, the same neurons were identified by reference to a grid scored on the bottom of the dishes before culture. Accurate drawings of these neurons were made at two-hourly intervals using a drawing tube attached to an inverted phase-contrast microscope. To maintain the temperature at 37 °C during the period of observation, the microscope stage was enclosed in an incubator chamber. The lengths of serially drawn neurons were measured by tracing their outline onto a digitizing table linked to a computer running a graphics program.

targets and become dependent on target-derived trophic factors for survival and growth. Thus, although this work has provided valuable information on the cell biology of neurite elongation, the relevance of the observed growth rates, which ranged from 10 to 250 $\mu\text{m h}^{-1}$, to the developmentally appropriate stages of axonal growth *in vivo*, is questionable.

The work presented here raises the question of when growth rate is regulated in sensory neurogenesis. This may be determined either in pre-migratory neuron progenitor cells in the corresponding ectodermal placodes, in progenitor cells after they have reached the developing ganglia or in the earliest stages of the differentiation of neurons from their progenitors. Elucidation of the stage at which growth rate is determined will provide some indication of the type of extracellular signals involved in its regulation. Furthermore, studies in early ganglia of the level of expression of genes encoding cytoskeletal proteins may provide some indication of the molecular basis of differential growth control.

I thank Peter Hollenbeck for the anti-neurofilament antiserum and Paul Martin for the Protargol. This work was supported by a grant from the Wellcome Trust.

Received 27 October 1988; accepted 9 January 1989.

1. Davies, A. M., Bandtlow, C., Heuman, R., Korsching, S., Rohrer, H. & Thoenen, H. *Nature* **326**, 353–358 (1987).
2. Speidel, C. C. *Havey Lect.* **36**, 126–158 (1941).
3. Jacobson, M. & Huang, S. *Dev Biol.* **110**, 102–113 (1985).
4. Davies, A. M. *Development* **100**, 307–311 (1987).
5. D'Amico-Martel, A. *Am. J. Anat.* **163**, 351–372 (1982).
6. D'Amico-Martel, A. & Noden, D. M. *Am. J. Anat.* **166**, 445–468 (1983).
7. Davies, A. M. *Trends Genet.* **4**, 139–143 (1988).
8. Ernsberger, U. & Rohrer, H. *Dev Biol.* **126**, 420–432 (1988).
9. Davies, A. M., Thoenen, H. & Barde, Y. A. *J. Neurosci.* **6**, 1897–1904 (1986).
10. Hughes, A. J. *Anat.* **87**, 150–161 (1953).
11. Nakai, J. *Am. J. Anat.* **99**, 81–130 (1956).
12. Bray, D. J. *Cell Biol.* **56**, 702–712 (1973).
13. Letourneau, P. *Dev Biol.* **44**, 77–91 (1975).
14. Lewis, J. H. *Zoon.* **6**, 175–179 (1978).

Arachidonic acid metabolites as intracellular modulators of the G protein-gated cardiac K^+ channel

Yoshihisa Kurachi*, Hiroyuki Ito*, Tsuneaki Sugimoto*, Takao Shimizu†, Ichiro Miki†§ & Michio Ui‡

* The 2nd Department of Internal Medicine, and † Department of Physiological Chemistry and Nutrition, Faculty of Medicine and ‡ Department of Physiological Chemistry, Faculty of Pharmaceutical Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

Arachidonic acid is released from cell membranes in response to receptor-dependent as well as receptor-independent stimulation in various cells, including cardiac myocytes^{1,2}. Arachidonic acid is converted to prostaglandins by cyclooxygenase and to leukotrienes by 5-lipoxygenase¹, metabolites which are very biologically active and modulate cellular functions such as platelet aggregation, smooth muscle contraction and neural excitation. The molecular mechanisms underlying their modulations are, however, still badly understood. Here, we report that the 5-lipoxygenase metabolites of arachidonic acid activate the pertussis toxin-sensitive G protein-gated muscarinic K^+ channel ($\text{I}_{\text{K-Ach}}$): arachidonic acid activation of $\text{I}_{\text{K-Ach}}$ was prevented by the lipoxygenase inhibitors, nordihydroguaiaretic acid and AA-861; leukotriene A_4 and C_4 activated $\text{I}_{\text{K-Ach}}$. The activation occurred in pertussis toxin-treated atrial cells and ceased when inside-out patches were formed but the patches were still susceptible to stimulation by GTP and to inhibition by GDP- β -S. These results indicate that arachidonic acid metabolites may stimulate the G-protein in a receptor-independent way.

§ Permanent address: Kyowa Hakko Kogyo Co., Pharmaceutical Research Laboratories, Nagaizumi-cho, Sunto-gun, Shizuoka 411, Japan.

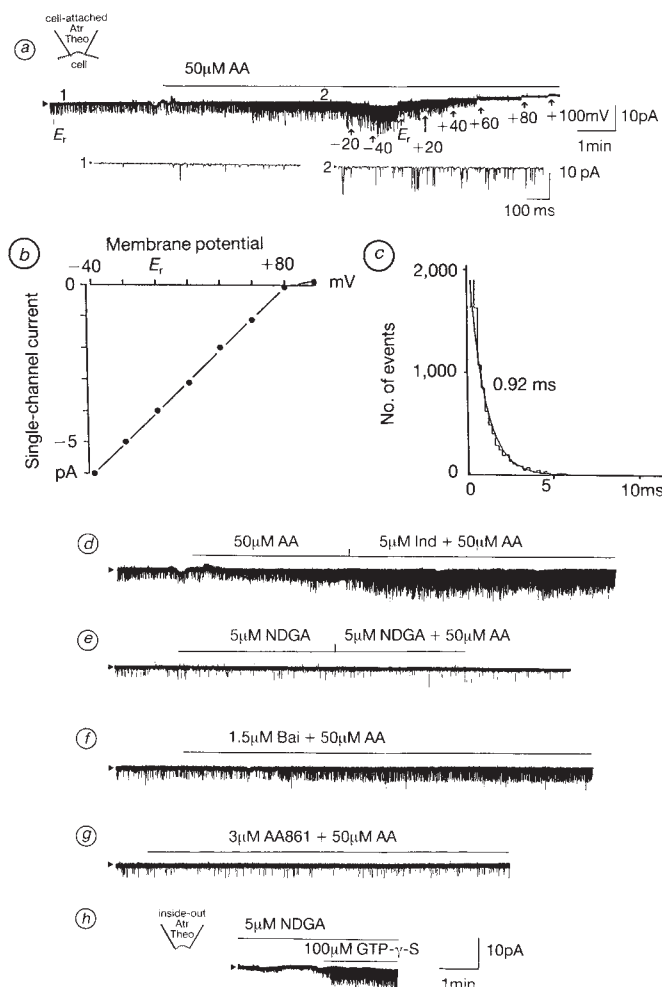


Fig. 1 Arachidonic acid activation of the muscarinic K^+ channel and the effects of inhibitors of the lipoxygenase and cyclooxygenase pathways. In the cell-attached patch, when arachidonic acid (AA, 50 μM , Cayman Biochemicals) was applied to the bath, it activated a K^+ channel (a). The membrane potential of the patch is indicated below the current trace. The resting potential (E_r) of the cell in this case was -79 mV. The locations of the expanded current traces are indicated (1, 2) in the upper trace. b, Current-voltage relation and c, open-time histogram of the arachidonic-acid-induced K^+ channel at E_r . d, Indometacin (Ind, 5 μM) enhanced arachidonic acid activation of $\text{I}_{\text{K-Ach}}$. Although arachidonic acid (50 μM) activated $\text{I}_{\text{K-Ach}}$ in 15/24 control cells (54%), it activated the channel in all cells ($n = 14$) pre-treated with indometacin. e, g, NDGA (5 μM) and AA-861 (3 μM) blocked activation of $\text{I}_{\text{K-Ach}}$. f, Baicalein (Bai, 1.5 μM) did not affect activation. These inhibitors (indometacin, NDGA, baicalein and AA-861) suppressed neither the ACh or the adenosine-induced activation of $\text{I}_{\text{K-Ach}}$ in the cell-attached patches nor intracellular GTP- γ -S-induced activation of $\text{I}_{\text{K-Ach}}$ in the inside-out patches at these concentrations. (h) Atr, atropine, Theo, theophylline.

Methods. Single atrial cells were dispersed from guinea-pig heart, using collagenase⁵. All experiments were performed at 32–36 °C. The patch-clamp technique was used to record currents flowing through $\text{I}_{\text{K-Ach}}$ in cell-attached and inside-out patch modes¹³. The pipette solution contained 145 mM KCl, 1 mM CaCl_2 , 1 mM MgCl_2 , 5 mM HEPES-KOH (pH 7.4), 10 μM atropine and 100 μM theophylline. The bathing solution contained 136.5 mM NaCl, 5.4 mM KCl, 1.8 mM CaCl_2 , 0.53 mM MgCl_2 , 5.5 mM glucose, 5.5 mM HEPES-NaOH (pH 7.4). The internal solution was 140 mM KCl, 0.5 mM MgCl_2 , 5 mM EGTA, 5 mM HEPES-KOH (pH 7.3). GTP- γ -S was purchased from Boehringer. Concentrated stock solutions of inhibitors in ethanol were prepared fresh each day and were kept at 4 °C. Portions of the solutions were dissolved in the bathing solution. Inhibitors were purchased from Sigma. AA-861 (2, 3, 5-trimethyl-6-(12-hydroxy-5, 10-dodecadienyl)-1,4-benzoquinone) was a gift from Takeda (Osaka, Japan). The final concentration of ethanol was less than 0.5%, which did not affect the arachidonic acid-activation of $\text{I}_{\text{K-Ach}}$.

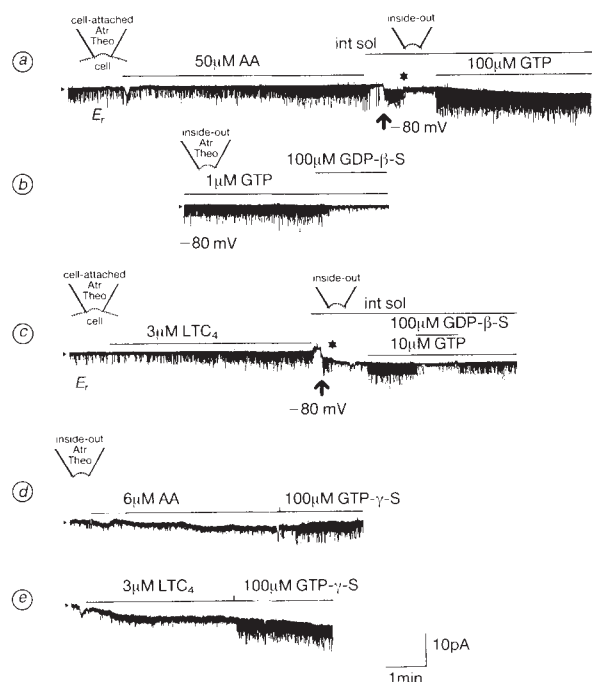


Fig. 2 Effects of exogenous arachidonic acid metabolites on I_{K-ACH} . 12-HPETE (40 μ M, *a*), 5-HPETE (4 μ M, *b*), LTA_4 (4.5 μ M, *c*) and LTC_4 (1.6 μ M, *d, e*) were added to the bath solution in the cell-attached patches. 5-HPETE, LTA_4 and LTC_4 activated I_{K-ACH} . LTC_4 (1.6–7.5 μ M) activated I_{K-ACH} even with perfusion of AA-861 (3 μ M, *e*).

Methods. LTA_4 methyl ester and LTC_4 were purchased from Ultrafine Chemicals (Manchester, UK) and Cayman Biochemicals (Michigan, USA), respectively. 12-HPETE, 12-HETE and 5-HPETE were prepared as previously described^{14,15}. LTB_4 was a gift from Ono Pharmaceutical (Osaka, Japan). All metabolites were kept at -80°C . Immediately before an experiment, a sample of the stock solution was added to the bathing solution and the mixture was sonicated. Final ethanol concentration was less than 0.5%, which did not affect arachidonic acid-activation of I_{K-ACH} .

Figure 1a shows that arachidonic acid activates a K^+ channel in atrial myocytes. When arachidonic acid was applied to the bath, the channel openings gradually increased after a delay of 1–2 min and reached a steady level within 10 min in the cell-attached patch. The pipette solution contained atropine and theophylline to block muscarinic and adenosine receptors respectively, both of which are linked to I_{K-ACH} by a pertussis-toxin-sensitive G-protein (G_K)^{3–5}. The steady-state channel activity, Np_0 (where N is the number of channels in the patch and p_0 is the probability that a channel is open) increased from 3.9×10^{-4} (Np_0 of the background activity, $Np_{0,back}$) to 2.773×10^{-2} in Fig. 1a. In seven experiments, Np_0 increased in the presence of arachidonic acid (50 μ M) by a factor of 58.9 ± 21.3 (mean $\pm$ s.d.). This level of channel opening persisted for more than 10 minutes after the arachidonic acid was washed out. The arachidonic acid-activated K^+ channel has the same conductance and kinetic properties as I_{K-ACH} ^{5,6}. The channel has strong inward rectification with a conductance of 45–50 pS in 150 mM K^+ in the pipette solution (cell-attached form) (Fig. 1b). The open-time histogram of the channel fits a single exponential curve with a time constant of about 1 ms (Fig. 1c).

Arachidonic acid is converted to various metabolites by cyclooxygenase and lipoxygenases¹ and may modulate ion channels⁷. A cyclooxygenase inhibitor, indometacin (5–10 μ M) enhanced rather than blocked arachidonic acid activation of I_{K-ACH} ($n = 14$, Fig. 1d). In contrast, nordihydroguaiaretic acid (NDGA, 5–10 μ M), a lipoxygenase inhibitor, prevented activation of I_{K-ACH} ($n = 7$, Fig. 1e), suggesting that lipoxygenase metabolites stimulate I_{K-ACH} but cyclooxygenase metabolites do not. We

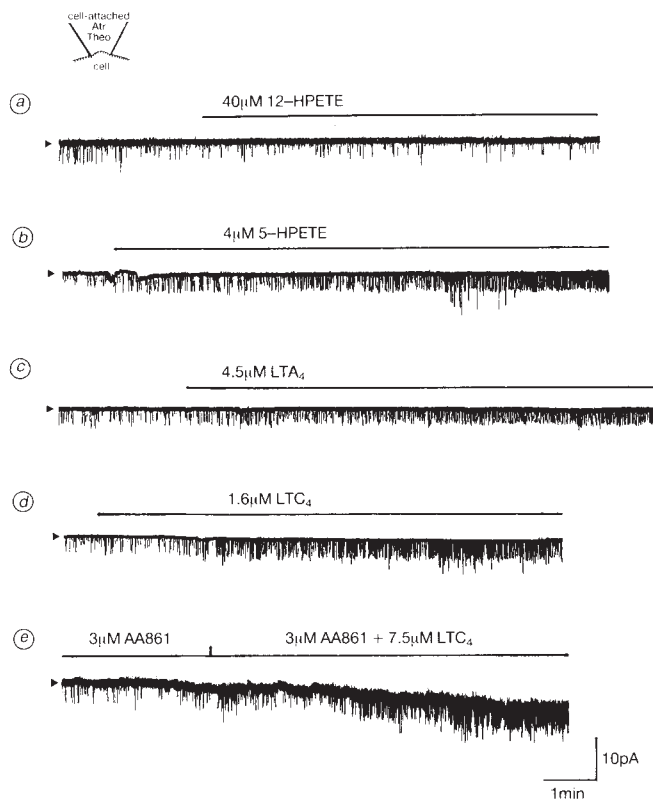


Fig. 3 Arachidonic acid-metabolite activation of I_{K-ACH} may involve a GTP-dependent process. After activation of I_{K-ACH} by arachidonic acid (50 μ M, *a*) or LTC_4 (3 μ M, *c*) reached a steady level, the patches were excised in the internal solution, yielding 'inside-out' patches (stars). GTP (10 or 100 μ M) was added to the internal solution as indicated by a bar in each current trace. The channel openings disappeared quickly in the inside-out patch but reappeared after application of GTP in both cases. After GTP was washed out from the bath, the channel openings faded again (data not shown). In the patches without pretreatment with arachidonic acid or LTC_4 , intracellular GTP did not open I_{K-ACH} ($n = 5$). The channel openings induced by intracellular GTP (1 or 10 μ M in these examples) were inhibited by GDP- β -S (100 μ M, *b, c*). In the inside-out patches, arachidonic acid (6–10 μ M) and LTC_4 (1.5–3 μ M), applied to the internal solution, did not activate I_{K-ACH} , whereas GTP- γ -S (100 μ M) activated the channel in the same patches (*d, e*). As arachidonic acid ($>10 \mu$ M) easily broke the patch membrane, we could not examine the effects of high concentrations of arachidonic acid on I_{K-ACH} in the inside-out patches. The cells in *b* and *c* were pre-treated with pertussis toxin (100 ng ml⁻¹) for 6 hours at 30°C . In these cells, ACh and adenosine could not activate I_{K-ACH} but arachidonic acid (50 μ M, *b*) and LTC_4 (1.5–3 μ M, *c*), added in the bath solution, gradually activated I_{K-ACH} in the cell-attached patches.

Methods. GTP and GDP- β -S were purchased from Sigma and Boehringer respectively. Pertussis toxin was from Funakoshi (Tokyo, Japan).

examined the effects of baicalein (1.5–3 μ M), an inhibitor of 12-lipoxygenase⁸, and AA-861 (3 μ M), an inhibitor of 5-lipoxygenase⁹. Arachidonic acid increased Np_0 to 50.2 ± 15.6 ($n = 5$) times $Np_{0,back}$ in the baicalein-treated cells (Fig. 1f), but did not activate I_{K-ACH} in AA-861-treated cells (Fig. 1g, $n = 7$). NDGA and AA-861 did not affect the GTP- γ -S-activation of I_{K-ACH} (Fig. 1h).

Neither 12-hydroperoxyeicosatetraenoic acid (12-HPETE, 8–40 μ M) nor 12-hydroxyeicosatetraenoic acid (12-HETE, 15 μ M) activated I_{K-ACH} consistent with the inhibitor results (Fig. 2a, $n = 10$). A 5-lipoxygenase metabolite, 5-HETE (15 μ M) did not increase channel opening either (data not shown, $n = 5$). In contrast, 5-HPETE (4 μ M) and its metabolites, leukotriene A_4 (LTA_4 , 4.5 μ M) and LTC_4 (1.6 μ M) activated I_{K-ACH} (Fig.

2b-d, $n = 4$). LTC₄ (1.6 μ M) increased Np_0 by a factor of 57.2 ± 24.6 ($n = 4$) over $Np_{0,back}$. Neither AA-861 ($n = 3$) nor NDGA ($n = 2$) prevented LTC₄ activation of I_{K-ACH} (Fig. 2e). Another compound derived from LTA₄, LTB₄ (6 μ M), did not activate I_{K-ACH} (data not shown, $n = 3$). These results indicate that 5-lipoxygenase metabolites, probably LTC₄ and/or related metabolites¹⁰, activate I_{K-ACH} .

I_{K-ACH} activated by arachidonic acid and LTC₄ disappeared quickly in the inside-out patches but reappeared on application of GTP (Fig. 3a, $n = 7$; c, $n = 5$), although this was inhibited by GDP- β -S (Fig. 3b, c). When arachidonic acid (1–10 μ M, $n = 5$) or LTC₄ (3 μ M, $n = 5$) was applied to inside-out patches, I_{K-ACH} was not activated, although GTP- γ -S activated I_{K-ACH} in the same patches (Fig. 3d, e). As the pipette solution contained atropine and theophylline, muscarinic and adenosine receptors cannot be involved in this process. In addition, arachidonic acid and LTC₄ activated I_{K-ACH} in pertussis toxin-treated atrial cells (Fig. 3b, c), where G_K is uncoupled from receptors^{3,5,11,12}.

Received 31 October 1988; accepted 4 January 1989.

1. Needleman, P., Turk, J., Jakschik, B. A., Morrison, A. R. & Lefkowitz, J. B. *A. Rev. Biochem.* **55**, 69–102 (1986).
2. Axelrod, J., Burch, R. M. & Jelsema, C. L. *Trends Neurosci.* **11**, 117–123 (1988).
3. Pfaffinger, P. J., Martin, J. M., Hunter, D. D., Nathanson, N. M. & Hille, B. *Nature* **317**, 536–538 (1985).
4. Breitwieser, G. E. & Szabo, G. *Nature* **317**, 538–540 (1985).
5. Kurachi, Y., Nakajima, T. & Sugimoto, T. *Pflügers Arch. ges. Physiol.* **407**, 264–274 (1986).
6. Logothetis, D. E., Kurachi, Y., Galper, J., Neer, E. J. & Clapham, D. E. *Nature* **325**, 321–326 (1987).

Our observations suggest that these substances may stimulate GDP/GTP exchange of G_K in a receptor-independent way, or that they may bind to membrane receptors coupled to I_{K-ACH} , for example through pertussis toxin-insensitive G-proteins. When arachidonic acid and its metabolites (LTA₄ and C₄) were added to the pipette solution, there was no activation of I_{K-ACH} within 2–5 min (data not shown), perhaps favouring the former possibility.

As arachidonic acid is released by stimulation by various agents^{1,2}, it is possible that arachidonic acid metabolites are common intracellular messengers for the G-protein-gated K^+ channel in physiological and pathophysiological regulation of cardiac excitation.

We are grateful to Dr Yoshio Maruyama for his kind comments. This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan, and the Research Program on Cell Calcium Signalling in the Cardiovascular System.

7. Piomelli, D. *et al. Nature* **328**, 38–43 (1987).
8. Sekiya, K. & Okuda, H. *Biochem. biophys. Res. Commun.* **105**, 1090–1095 (1982).
9. Yoshimoto, T. *et al. Biochim. biophys. Acta* **713**, 470–473 (1982).
10. Shimizu, T. *Int. J. Biochem.* **20**, 661–666 (1988).
11. Katada, T. & Ui, M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3129–3133 (1982).
12. Florio, V. A. & Sternweis, P. C. *J. Biol. Chem.* **260**, 3477–3483 (1985).
13. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85–100 (1981).
14. Kitamura, S. *et al. Eur. J. Biochem.* **176**, 725–731 (1988).
15. Shimizu, T., Rådmark, O. & Samuelsson, B. *Proc. natn. Acad. Sci. U.S.A.* **81**, 689–693 (1986).

G-protein $\beta\gamma$ -subunits activate the cardiac muscarinic K^+ -channel via phospholipase A₂

Donghee Kim*, Deborah L. Lewis*, Linda Graziadei‡, Eva J. Neer†, Dafna Bar-Sagi§ & David E. Clapham*||

* Departments of Pharmacology, Physiology, and Biophysics, Mayo Foundation, Rochester, Minnesota 55905, USA

† Cardiovascular Division, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

‡ Cellular and Developmental Biology Program, SUNY, Stony Brook, New York, USA

§ Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA

Muscarinic receptors of cardiac pacemaker and atrial cells are linked to a potassium channel (I_{K-ACH}) by a pertussis toxin-sensitive GTP-binding protein^{1,2,3}. The dissociation of G-proteins leads to the generation of two potential transducing elements, α -GTP and $\beta\gamma$ ^{4–6}. I_{K-ACH} is activated by G-protein α - and $\beta\gamma$ -subunits applied to the intracellular surface of inside-out patches of membrane^{7–10}. $\beta\gamma$ has been shown to activate the membrane-bound enzyme phospholipase A₂ in retinal rods¹¹. Arachidonic acid, which is produced from the action of phospholipase A₂ on phospholipids, is metabolized to compounds which may act as second messengers regulating ion channels in *Aplysia*¹². Muscarinic receptor activation leads to the generation of arachidonic acid in some cell lines¹³. We therefore tested the hypothesis that $\beta\gamma$ activates I_{K-ACH} by stimulation of phospholipase A₂. When patches were first incubated with antibody that blocks phospholipase A₂ activity¹⁴, or with the lipoygenase inhibitor, nordihydroguaiaretic acid, $\beta\gamma$ failed to activate I_{K-ACH} . Arachidonic acid and several of its metabolites derived from the 5-lipoxygenase pathway, activated the channel. Blockade of the cyclooxygenase pathway did not inhibit arachidonic acid-induced channel activation. We conclude that the $\beta\gamma$ -subunit of G-proteins activates I_{K-ACH} by stimulating the production of lipoygenase-derived second messengers.

$\beta\gamma$ applied to inside-out patches of neonatal rat atrial membranes activated a 35 ± 5 pS channel with a mean open time of ~ 1 ms. $\beta\gamma$ began to activate I_{K-ACH} at 200 pM⁹. Activation occurred consistently using 1–10 nM $\beta\gamma$ (118/123 patches in chick and rat atria, ref. 9; or 12/12 using 2 nM $\beta\gamma$ in the present study; Fig. 1a). Preincubation of inside-out rat atrial patches with affinity-purified antibody to phospholipase A₂ (PLA₂) completely blocked activation by 2 nM $\beta\gamma$ (5/5 patches; Fig. 1b). This same antibody has been shown to inhibit PLA₂-mediated production of lysophosphoglycerides in membrane preparations of rat fibroblasts¹⁴. Anti-PLA₂ antibody did not block subsequent activation by GTP- γ -S (5/5; Fig. 1b) or by GTP when acetylcholine was present in the pipette (3/3), suggesting that activation can also occur via α -subunits or α -subunit-activated intermediates. Boiled anti-PLA₂ antibody (3/3; Fig. 1c) and control IgG prepared from preimmune serum (2/2; Fig. 1d) did not block $\beta\gamma$ -dependent activation. Incubation of the patch with antibody for periods of less than five minutes was ineffective in blocking K^+ -channel activation by $\beta\gamma$. The possibility that PLA₂ contaminants were the cause of the $\beta\gamma$ activation was eliminated by two controls. First, exogenous bee venom PLA₂ (10–100 units ml⁻¹; Sigma) did not activate channels. Second, anti-PLA₂ antibody preincubated with $\beta\gamma$ for 30 minutes before it was applied to the patch (fivefold molar excess of antibody) did not block activation by 2 nM $\beta\gamma$ (4/4). As arachidonic acid is generated from phospholipids by PLA₂, we tested several compounds and inhibitors of the lipoygenase and cyclooxygenase pathways directly on atrial patches. Nordihydroguaiaretic acid (NDGA) prevented $\beta\gamma$ (2 nM)-induced activation at 1 μ M (5/5 patches) and 10 μ M (8/8 patches; Fig. 1e).

Bath application of arachidonic acid (50–100 μ M) activated I_{K-ACH} in cell-attached patches within several minutes (Fig. 2a). Lower concentrations (10 μ M) were sufficient for the activation of inside-out patches (Fig. 2b). Inhibition of both the lipoygenase and cyclooxygenase pathways by eicosatetraynoic acid, or the lipoygenase pathway by NDGA, blocked arachidonic acid-mediated activation of I_{K-ACH} in both cell-attached (Fig. 2c, d) and inside-out patches (data not shown). Blockade of the cyclooxygenase pathway by cell and patch incubation with 20 μ M indometacin did not inhibit the arachidonic acid-induced channel activity, suggesting that prostaglandins are not activators of the channel. In cell-attached patches, 5-hydro-

|| To whom correspondence should be addressed.

Fig. 1 Antibody to PLA₂ inhibits $\beta\gamma$ activation of $I_{K_{ACH}}$. *a*, 2 nM $\beta\gamma$ activates a 40 pS K⁺-channel ($I_{K_{ACH}}$) in an inside-out patch from neonatal rat atria. Membrane potential held at -80 mV. *b*, An inside-out patch was held in a bath solution (see below) containing 100 nM anti-PLA₂ antibody for 20 min. When 2 nM $\beta\gamma$ was subsequently applied to the patch, channels were not activated. GTP- γ -S applied 5 min later activated $I_{K_{ACH}}$. *c*, Boiled anti-PLA₂ antibody (100 nM) applied for 25 min to an inside-out patch does not prevent activation of the channel by 2 nM $\beta\gamma$. *d*, Preincubation of the patch with rabbit anti-rat IgG (200 nM) did not block $\beta\gamma$ activation. *e*, Preincubation of the patch with 1 μ M NDGA blocked $\beta\gamma$ (2 nM)-dependent activation but not GTP- γ -S-dependent stimulation. All channels we identify as $I_{K_{ACH}}$ had 35 ± 5 pS conductances in 140 mM symmetrical KCl, mean open times ranging from 1.0 to 1.5 ms, and displayed inward rectification in the presence of Mg^{2+} on the intracellular surface of the patch.

Methods. Hearts from 1-2-day-old newborn rats were enzymatically dissociated¹⁹, plated at low density on circular glass coverslips and cultured at 37 °C in 5% CO₂ for 12-24 hours before use. All experiments were performed at 22-24 °C. The pipette and bath solution contained 140 mM K⁺, 140 mM Cl⁻, 5 mM EGTA, 2 mM Mg^{2+} , and 10 mM HEPES (pH 7.2). In all experiments using an antibody, the antibody was first added to the bath before inside-out patch formation. After a 20-30 minute incubation of the inside-out patch with antibody, the patch pipette was brought into the mouth of polyethylene tubing from which $\beta\gamma$ flowed into the bath. Polyclonal antibodies to porcine pancreatic PLA₂ were purified by affinity chromatography using immobilized PLA₂ as previously described¹⁴. This affinity-purified PLA₂ antibody detected a single polypeptide of relative molecular mass 16,000 on immunoblots of total cellular protein from rat fibroblasts and inhibited the activity of membrane bound PLA₂¹⁴. Control IgG was prepared from rabbit preimmune serum. Data were recorded using a List EPC-7 patch clamp amplifier. Raw data were plotted with a Gould thermal array recorder (cut-off frequency 2000 Hz). Other recording details and calculation of Np_0 for single channel analysis have been previously described⁷. The relative increase in channel activity (Np_0) over basal levels is shown in each panel ($\times$).

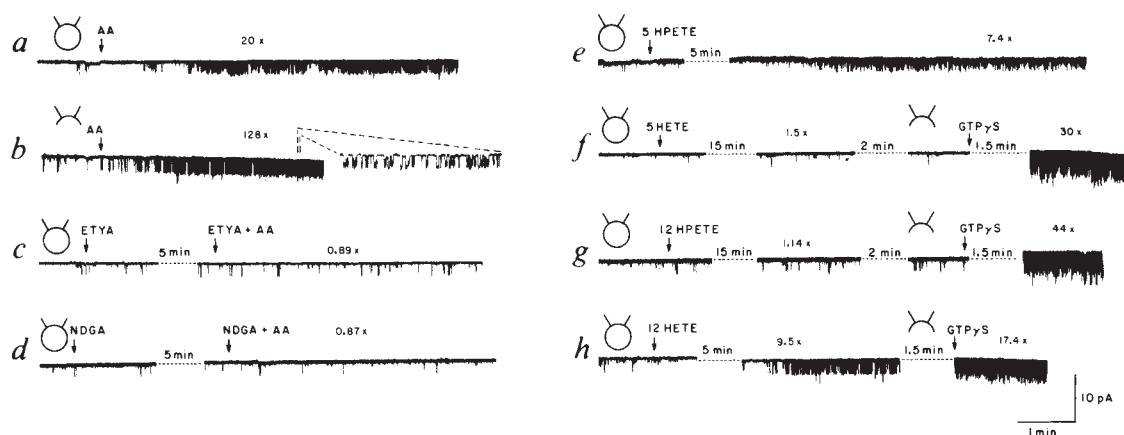
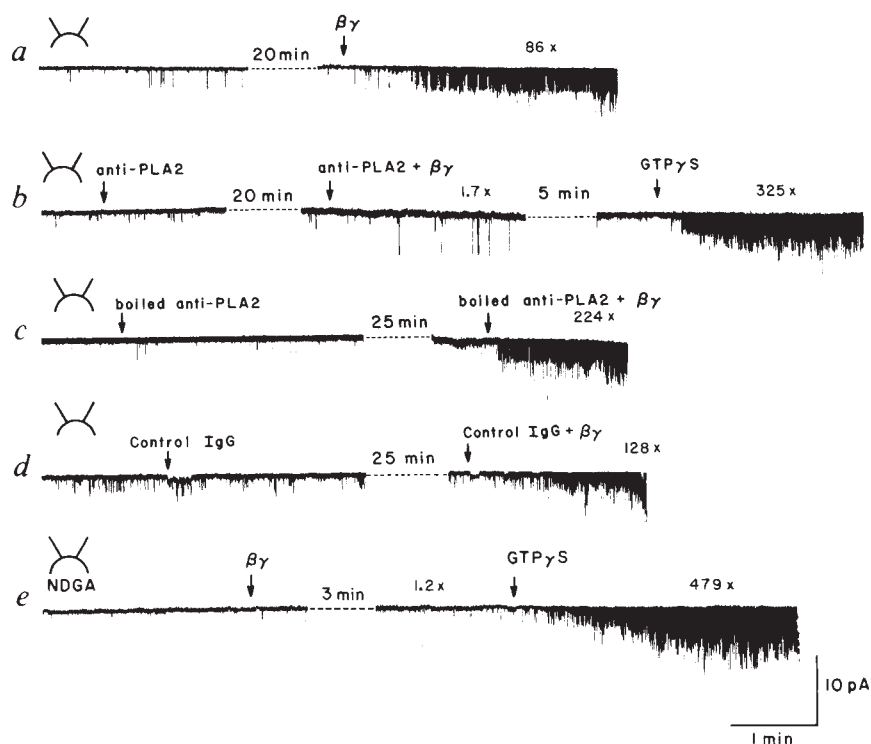


Fig. 2 Arachidonic acid and lipoxygenase derivatives activate $I_{K_{ACH}}$. Arachidonic acid (AA) activated the channel when applied to *a*, cell-attached patches (50 μ M) or *b*, inside-out patches (10 μ M). A single channel recording at higher time resolution is shown at the right. *c*, *d*, ETYA (10 μ M), which inhibits arachidonic acid metabolism to cyclooxygenase and lipoxygenase derivatives, and NDGA (20 μ M), which inhibits metabolism of arachidonic acid to lipoxygenase derivatives both block arachidonic acid-(50 μ M) induced activation of the K⁺-channel. *e-h*, Arachidonic-acid derived metabolites of the lipoxygenase pathway, 5-HPETE and 12-HETE (*e, h*), activated the channel (100 μ M; applied for 5 min). 5-HETE, 12-HPETE (*f, g*), and 15-HETE (data not shown) did not activate $I_{K_{ACH}}$ in cell-attached (100 μ M; applied for >15 min) and inside-out patches (10 μ M; data not shown) whereas 10 μ M GTP- γ -S immediately activated the channel in the same patches. Note that 12-HETE did not maximally stimulate the channel as shown by the subsequent GTP- γ -S activation.

Methods. Arachidonic acid (Calbiochem) and its metabolites (Cayman and generous gifts from Dr Rod Matthews, Upjohn Pharmaceuticals, and Dr Jeff Stadel; Smith, Kline, and French, Inc.) were always kept under nitrogen or argon gas and used immediately after suspension. Ethanol was evaporated and the compounds dissolved in the perfusion solution by sonication for 20 s. NDGA (Sigma) or ETYA (Biomol) was dissolved at 200 mM concentrations in dimethylsulphoxide and applied to membrane patches at a final concentration of 10 μ M. Dimethylsulphoxide (0.025%) or ethanol (up to 1%) had no effect on channel activity in prior control experiments.

peroxyicosatetraenoic acid (5-HPETE), a metabolite of arachidonic acid produced by the 5-lipoxygenase pathway, activated $I_{K_{ACH}}$ in three out of five patches (Fig. 2*e*). Notably, 12-hydroxyicosatetraenoic acid (12-HETE) also occasionally activated the channel (2/7 patches) whereas 5-HETE, 12-

HPETE, and 15-HETE did not (50-100 μ M; tested up to 30 min, $n = 4$ for each case; Fig. 2*f, g*). NDGA (10 μ M) did not block the activation by 5-HPETE (2/5 cells in NDGA were activated by 5-HPETE compared to 3/5 controls).

Figure 3 summarizes the action of leukotrienes on $I_{K_{ACH}}$.

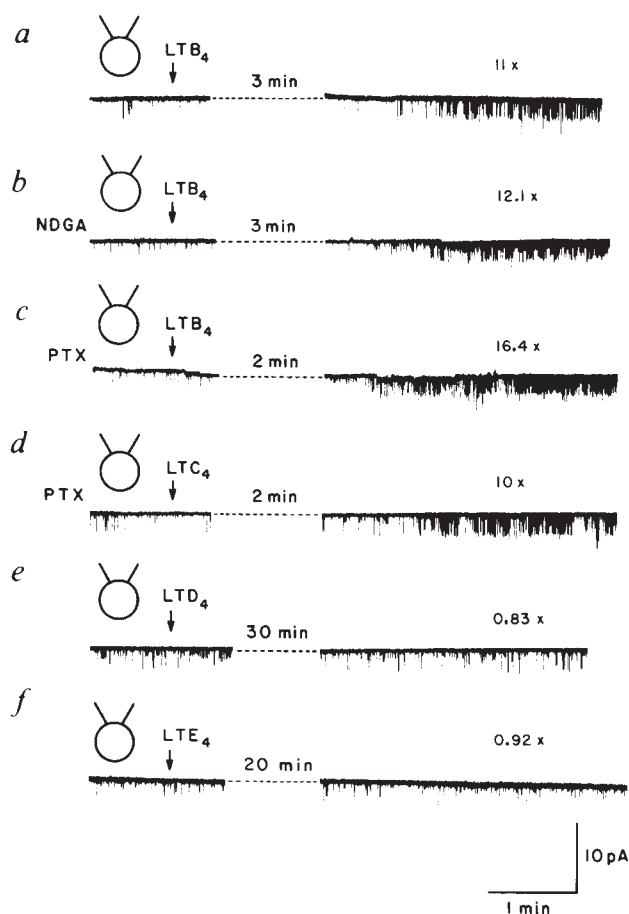


Fig. 3 *a*, Leukotriene B₄ (10 μM) activated I_{K,ACH} in cell-attached patches of atrial membrane, even in the presence of 10 μM NDGA (*b*). *c*, *d*, Pertussis toxin pretreatment (100 ng ml⁻¹ for 12 hours) of cells did not affect activation by either leukotriene B₄ or C₄. Previous work has shown that pertussis toxin blocks muscarinic activation of the channel^{1,3,7}. *e*, *f*, Leukotrienes D₄ and E₄ (10 μM) did not activate the channel.

Leukotrienes are products of the 5-lipoxygenase pathway and may act through specific receptors¹⁵. Both leukotriene B₄ (LTB₄) (3/3) and C₄ (3/3) activated the K⁺ channel whereas the D₄ and E₄ forms did not. Pertussis toxin pretreatment did not block the effects of leukotrienes (Fig. 3). (In control experiments, acetylcholine activation was always blocked by pertussis toxin treatment; ref. 7.) NDGA (10 μM) did not block LTB₄-induced activation (6/9 inside-out patches). At present, we cannot distinguish between direct effects of the metabolites on the channel, binding of leukotrienes to membrane receptors and activation of the K⁺ channel through an unknown cascade, or destabilization of G-proteins linked to the muscarinic receptor (for example, releasing an activated α-subunit). A purified channel preparation would be needed to test the first possibility; the second requires specific leukotriene antagonists not yet available. The other alternative, destabilization of a G-protein is not supported by the available evidence. First, leukotrienes activated the channel in pertussis toxin-treated patches. Addition of GTP to patches activated first by arachidonic acid did not quantitatively increase the activity. Arachidonic acid, or its metabolites, were most effective, however, in cell-attached patches (13/65, or 20% of inside-out patches responded to arachidonic acid compared to more than 90% of cell-attached patches), indicating the importance of intracellular components or local GTP levels. A second issue is the effectiveness of arachidonic acid and its derivatives in stimulating the channel compared to βγ or GTP-γ-S. On average, βγ increased $\overline{Np_0}$ (where N is the number of channels and p_0 is the probability of opening) 180-fold; GTP-γ-S, 175-fold; 5-HPETE, 11-fold; and LTB₄, 30-fold. Thus,

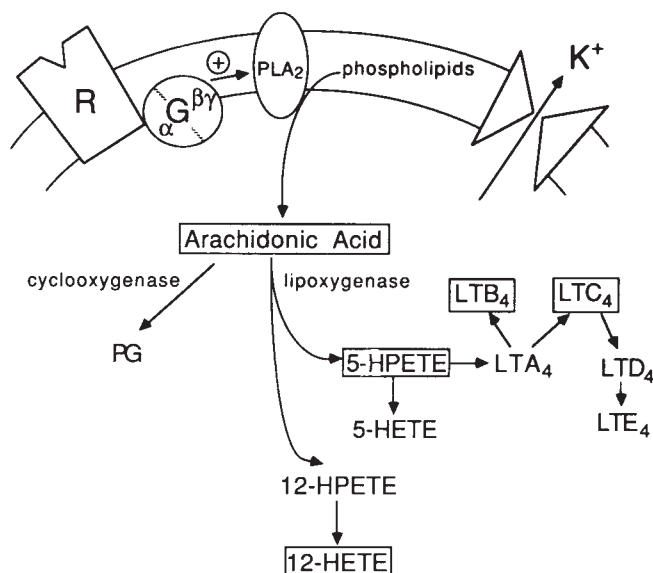


Fig. 4 Proposed mechanism of the G-protein βγ-subunit activation of the cardiac muscarinic K⁺-channel. βγ, released from the G-protein bound to the muscarinic receptor or from another G-protein-linked receptor, activates PLA₂ and lipoxygenase-derived compounds are released. Boxed metabolites/compounds activated the channel when applied to the patch. As these compounds are unstable, we do not know if 5-HPETE, 12-HETE, and the leukotrienes B₄ and C₄ all are direct mediators or whether the leukotrienes and distal metabolites are activators. Activation may occur through destabilization and release of α from the G-protein αβγ-heterotrimer, direct activation of the channel, or leukotriene receptor binding.

although it is clear that βγ activates the channel by PLA₂, products of this pathway added exogenously do not reproduce the full activation induced by intrinsic second messengers, suggesting that the local organization of enzymes and effectors is important.

To test the possibility that arachidonic acid was generated from phosphatidyl inositol, not by PLA₂ but by the sequential actions of phospholipase C and diacylglycerol lipase¹⁶, we pre-treated cells with 100 μM neomycin, an inhibitor of phospholipase C. Neomycin did not inhibit βγ-mediated activation of the channel ($n=3$). It is unlikely that protein kinase C stimulated the channel; channel activity was unaffected by exposure to the phorbol esters 12-*o*-tetradecanoylphorbol-13-acetate and 1-oleoyl-2-acetyl-glycerol in cell-attached and inside-out patches (D. E. Logothetis and D. E. Clapham, unpublished data). Also, as calcium was buffered to the 10 nM range using EGTA, the PLA₂ enzyme in our system would have to be relatively insensitive to calcium concentrations as has been shown for some cardiac PLA₂ enzymes¹⁷.

Block of PLA₂ by antibody, or of lipoxygenase products by NDGA, did not eliminate acetylcholine-induced activity (data not shown), suggesting that the α-subunit has an important, but not necessarily exclusive, role in muscarinic coupling to the ion channel. If the conclusions drawn from α-antibody experiments¹⁸ are correct, the α-subunit may be the main mechanism of muscarinic coupling to the channel. There is no evidence, however, for direct α-gating of I_{K,ACH} (rather than indirect activation by the α-subunit through intermediates) on the basis of simple protein/patch reconstitution.

Figure 4 summarizes the proposed mechanism for βγ activation of I_{K,ACH}. The final mechanism of channel activation through the arachidonic acid-dependent pathway is unknown but does not seem to involve a pertussis toxin-sensitive link to a receptor. Were it not for our demonstration of occasional activation by 12-HETE, one might assume that 5-lipoxygenase metabolites activate the channel because they generate leukotrienes as the final effectors. *In vivo* conversion of 12-

lipoxygenase products to the 5-lipoxygenase pathway is unlikely. Perhaps several lipoxygenase products activate the channel through a less specific mechanism than receptor coupling, such as release of the G-protein α -subunit or direct channel activation. A demonstration of the physiological significance of this pathway in heart will depend on finding receptors that increase arachidonic acid in conjunction with increased K^+ channel activity.

We thank Rose Goldsmith for culturing heart cells and Dr Yoshihisa Kurachi for informing us of his discovery that arachidonic acid activated $I_{K,ACH}^{20}$. This work was supported by NIH grants to D.K., E.J.N., D.B.-S. and D.E.C.

Received 31 October; accepted 16 December 1988.

1. Pfaffinger, P., Martin, J., Hunter, D., Nathanson, N. & Hille, B. *Nature* **317**, 536–538 (1985).
2. Breitwieser, G. & Szabo, G. *Nature* **317**, 538–540 (1985).
3. Kurachi, Y., Nakajima, T. & Sugimoto, T. *Pflügers Arch. ges. Physiol.* **407**, 264–274 (1986).
4. Neer, E. J. & Clapham, D. E. *Nature* **333**, 129–134 (1988).
5. Gilman, A. G. *A. Rev. Biochem.* **56**, 615–650 (1987).
6. Stryer, L. & Bourne, H. R. *A. Rev. cell Biol.* **2**, 391–419 (1986).
7. Logothetis, D. E., Kurachi, Y., Galper, J., Neer, E. J. & Clapham, D. E. *Nature* **325**, 321–326 (1987).
8. Codina, J., Yatani, A., Grenet, D., Brown, A. M. & Birnbaumer, L. *Science* **236**, 442–445 (1987).
9. Logothetis, D. E., Kim, D., Northup, J., Neer, E. J. & Clapham, D. E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5814–5818 (1988).
10. Cerbai, E., Klockner, U. & Isenberg, G. *Science* **240**, 1782–1783 (1988).
11. Jelsema, C. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3625–3627 (1987).
12. Piomelli, D. *et al. Nature* **328**, 38–43 (1987).
13. DeGeorge, J. J., Morell, P., McCarthy, K. D. & Lapetina, E. G. *J. biol. Chem.* **261**, 3428–3433 (1986).
14. Bar-Sagi, D., Suhan, J. P., McCormick, F. & Feramisco, J. R. *J. Cell Biol.* **106**, 1649–1658 (1988).
15. Parker, C. W. *A. Rev. Immun.* **5**, 65–84 (1987).
16. Irvine, R. F. *Biochem. J.* **204**, 3–16 (1982).
17. Loeb, L. A. & Gross, R. W. *J. biol. Chem.* **261**, 10467–10470 (1986).
18. Yatani, A. *et al. Science* **241**, 828–831 (1988).
19. Simpson, P. & Savion, S. *Circulation Res.* **50**, 101–116 (1982).
20. Kurachi, Y. *et al. Nature* **337**, 555–557 (1989).

Thymic cortical epithelial cells can present self-antigens *in vivo*

Robin G. Lorenz & Paul M. Allen

Department of Pathology, Washington University School of Medicine, 660 S. Euclid Avenue, St Louis, Missouri 63110, USA

Antigens present during neonatal life are recognized as self and individuals are tolerant to these antigens¹. In normal individuals T cells are tolerant to most self proteins but we still know little of the mechanism(s) by which tolerance is established. A requisite part of the current negative selection model of self tolerance is the expression of self proteins complexed with major histocompatibility complex molecules in the thymus^{2,3}. As MHC proteins bind antigens and present them to the receptor on the antigen-specific T cell, then for tolerance to self to occur, it is possible that each self protein must be processed and presented by an MHC molecule. As a result of the development of a unique T-cell hybrid reactive to the self protein murine haemoglobin, we have shown that in normal animals this self protein is continuously processed and potentially presented in an MHC-restricted manner⁴. Here we show that self haemoglobin is being processed and presented by thymic antigen-presenting cells as early as gestational day 14. We also demonstrate that three types of thymic stromal cells, namely macrophages, dendritic cells and cortical epithelial cells, can present the haemoglobin self antigen *in vivo*. This surprising presentation of a self antigen by thymic cortical epithelial cells implies that they could be involved in T-cell development and negative selection.

Using a unique murine T-cell hybridoma specific for an allelic form of murine haemoglobin (Hb), we have recently shown that the self protein Hb is constitutively processed and presented by thymic antigen-presenting cells (APCs) *in vivo*⁴. The presence

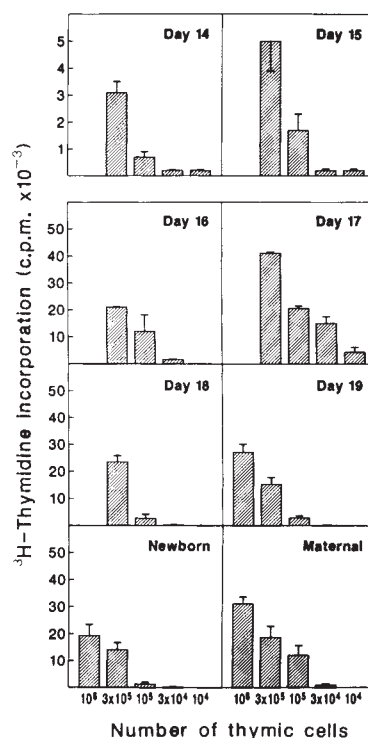
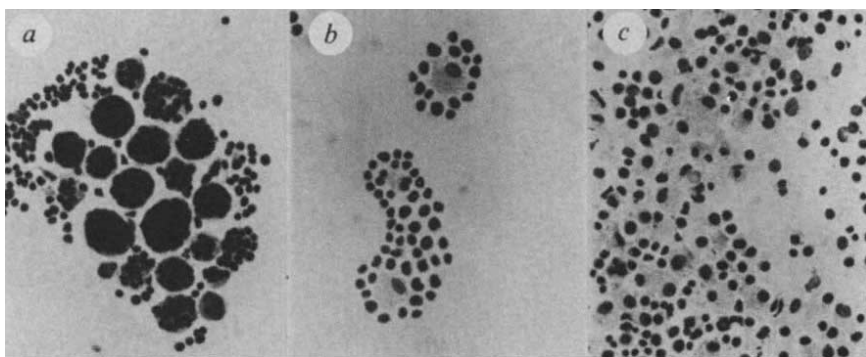


Fig. 1 Hb/1a complexes are detected by the YO1.6 hybrid on fetal thymic APCs. The stimulation assay was performed by mixing live thymic cell suspensions with the YO1.6 hybridoma (10^5 cells per well). No exogenous antigen was added. Stimulation was also detected if the thymic cell suspension was fixed with paraformaldehyde before the addition of the T-cell hybridoma (data not shown). T-cell stimulation was measured by quantitating the release of interleukin-2 (IL-2) by the YO1.6 hybridoma. 3 H-thymidine incorporation by the IL-2 dependent cell line CTLL-2 was used for lymphokine quantification¹⁹. Generation and characterization of the YO1.6 hybridoma is described in ref. 4. Briefly, a murine T-cell hybridoma specific for murine haemoglobin was generated by using two H-2^k mouse strains that express different alleles of Hb. The T-cell hybridoma YO1.6 was generated from CE/J mice (H-2^k, Hbb^s) and is specific for an allelic product of the murine Hb gene present in CBA/J mice (H-2^k, Hbb^d) (ref. 4). The YO1.6 hybrid recognizes the peptide Hb^{β^{minor}}(67–76) in association with the I-E^k MHC protein. We have ruled out the possibility that YO1.6 is recognizing a cross-reactive cell surface determinant, such as that from the minor lymphocyte stimulatory locus, by showing that (1) bone marrow Mφ from CBA/J mice (Hbb^d) do not stimulate the hybridoma, (2) when the Mls-positive strains CE/J and C58/J are used as sources of APCs, there is no stimulation of the hybridoma and (3) when APCs are isolated from the Mls-negative strain CBA/CaJ, the YO1.6 hybrid is still stimulated (data not shown, see ref. 20). In all cases, the YO1.6 hybridoma can be stimulated by these APC populations when exogenous Hb is added. Embryonic thymuses were obtained from matings between the inbred strain CBA/J (H-2^k, Hbb^d) from the Jackson Laboratory. Males and females were caged together overnight and checked for vaginal plugs the next morning. Day 0 is the day of finding a vaginal plug. It is known that the antigen-specific $\alpha\beta$ TCR is first expressed at gestational day 17 in fetal mice⁵, that class II MHC molecules are first detectable in the thymic epithelium at day 14 (ref. 21) and that the synthesis of adult type haemoglobin occurs at day 11 (ref. 22), and consequently all the components for self tolerance are present in murine fetuses by day 17. On the appropriate day the mice were sacrificed, the embryos dissected from the uterus, and the individual thymic lobes dissected free under a dissecting microscope. The thymic lobes were pooled, mechanically dissociated, and a single cell suspension prepared and added to a 96-well microtitre plate for the stimulation assay. Note the difference in scale of the vertical axis for days 14 and 15. Hbb represents the Hb allele expressed at the Hb β -chain locus and Hb^β stands for the Hb β -chain protein.

Fig. 2 Purified thymic APC populations: TNC (a), T-ROS (b) and M ϕ /DC (c). Photomicrographs were taken from cytocentrifuge preparations of the purified populations. Magnification, 400 $\times$. Staining procedures used haematoxylin and eosin. TNC are predominantly found in the outer cortical and subcapsular region and are MHC class I- and II-positive²³. They can first be isolated from fetal thymuses at day 17 of gestation²⁴. T-ROS have been detected as early as day 14 of gestation. Thymic DC are found exclusively in the thymic medulla and corticomedullary junction, and thymic M ϕ have been reported in the cortex, medulla and the corticomedullary junction. Preparation of purified populations of TNC (ref. 8), T-ROS (ref. 10) and M ϕ /DC (ref. 11) were essentially as described. Briefly, thymuses were removed from CBA/J mice (H-2^k, Hbb^d) when they were 5–8 weeks old. Individual thymuses in ice-cold RPMI-1640 buffered with 10 mM HEPES (pH 7.3) were trimmed free of any attached connective tissue or blood clots. The cleaned organs were minced with scissors and agitated in RPMI-HEPES for 10 min at room temperature. The supernatant containing most of the mechanically dissociable thymocytes was saved for use in the M ϕ /DC purification; the remaining tissue fragments were digested for two consecutive 15-min periods at room temperature with collagenase IV (Worthington; 0.5 mg ml⁻¹ in RPMI-HEPES, 0.5 ml per thymus). The supernatant from this digestion contains the T-ROS population, which was further purified by sequential 1 g fetal calf serum gradients⁸. The remaining tissue fragments were digested with 0.5 ml per thymus 0.25% trypsin (Boehringer) in Dulbecco's phosphate-buffered saline with DNase I at 100 μ g ml⁻¹ (Boehringer). Thymuses were digested for four consecutive 20-min periods at 37 $^{\circ}$ C. The supernatants from each round of digestion contain the TNC; they were pooled and the TNC further purified over fetal calf serum gradients. The M ϕ /DC population was purified by centrifugation over a discontinuous bovine serum albumin gradient. The APC population tested was the low-density fraction banding between 10 and 24% albumin.



of this self Hb/Ia complex in normal individuals indicates that the animals are tolerant to this complex, and indeed, the challenge of CBA/J mice with their own Hb does not elicit a response (data not shown). We have now used the Hb-specific hybridoma YO1.6 as a probe for the presence in fetal thymuses of the Hb/Ia complexes which are a test of this tolerance model. As shown in Fig. 1, low levels of self Hb/Ia complexes are detectable as early as gestational day 14, and levels equivalent to those in adult thymus are seen by day 16 and continue throughout fetal life. This thymic expression of self Hb/Ia complexes precedes that of the α TCR (T-cell antigen receptor) by 3 days⁵ and implies that Hb/Ia complexes are present at a time when the individual is learning tolerance to self.

At least three types of thymic stromal cells have been implicated in generating tolerance to self: cortical epithelial cells, medullary dendritic cells (DC) and macrophages (M ϕ) (ref. 6). Thymic nurse cells (TNC) can be isolated *in vitro* and are multicellular complexes formed between cortical epithelial cells and thymocytes (refs 7, 8; reviewed in ref. 9). The *in vitro* correlate of the medullary DC is the thymocyte rosette (T-ROS), which represents associations between thymocytes and bone marrow-derived class II positive dendritic-like cells or class II negative M ϕ (ref. 10). A second distinct population of thymic M ϕ can be isolated as a result of their low buoyant density which contains class II-positive M ϕ and DC (M ϕ /DC) (refs 11, 12).

If these three cell populations are involved in inducing tolerance to non-MHC self antigens, then they should have self protein/Ia complexes detectable on their surface. The photomicrographs in Fig. 2 are representative of the cell populations isolated and tested for stimulation of the YO1.6 hybridoma. It can be seen in Fig. 3 that all three distinct cell populations (TNC, T-ROS and M ϕ /DC) have detectable self Hb/Ia complexes on their surfaces. Our results contrast with the findings of Kyewski *et al.*, in which class II-positive cortical epithelial cells were inferred to be intrinsically deficient in antigen presentation¹³. As their assays used T-cell clones and were dependent on the diffusion of intravenously injected foreign antigen, it is likely that either the antigen did not diffuse to the cortical area containing the TNC, or that the TNC carried the antigen/MHC complexes but lacked some second signal needed for activation of the T-cell clones.

It is important to exclude the possibility that the Hb/Ia complexes detected by the YO1.6 hybrid are formed during the

TNC-isolation procedure, and we therefore prepared TNC from an equal mix of BALB/cJ (H-2^d, Hbb^d) and CE/J (H-2^k, Hbb^s) thymuses. If the complexes were correlates of *in vivo* complexes, then we should see no presentation to the YO1.6 hybridoma that is specific for Hb^{βd} in the context of the I-E^k MHC molecule, because the I-E^k molecule is not on the same APC population as the *in vivo*-processed Hb^{βd}. No presentation by the TNC was detected, indicating that the complexes previously identified were formed *in vivo* (data not shown). This population could present exogenous antigen added during the *in vitro* culture period. To demonstrate that the antigen-presenting capability of the TNC population was not a result of contamination by M ϕ or DC, we isolated the TNC population and character-

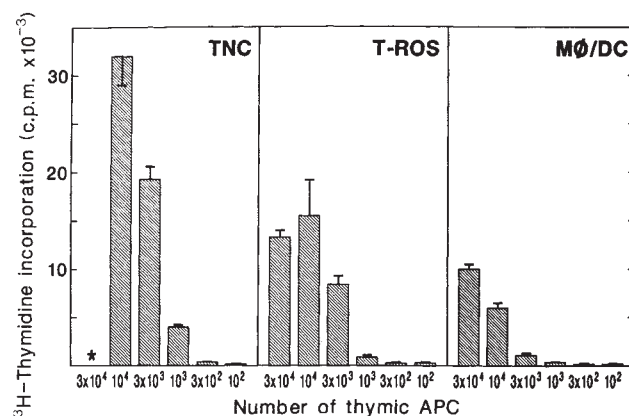


Fig. 3 Stimulation of the Hb-specific YO1.6 hybridoma by purified thymic APC populations. The stimulation assay is described in the legend to Fig. 1. This stimulation is blocked if the anti-I-E^k monoclonal antibody 14-4-4S is added to the *in vitro* culture (data not shown). Purified APC populations were obtained as described in Fig. 2. No exogenous antigen was added during the *in vitro* culture period. Comparison of six separate experiments revealed no statistically significant difference in presenting ability between the three APC populations (data not shown). Thus, contamination by a M ϕ or DC population cannot explain the level of presentation seen with TNC, as the contamination would have to approach 100% to explain the observed YO1.6 stimulation. It should be noted that our data do not prove that cortical epithelial cells can process antigen, only that they can present it *in vivo*. Further studies will be needed to determine the extent of their processing abilities. Asterisk indicates not tested.

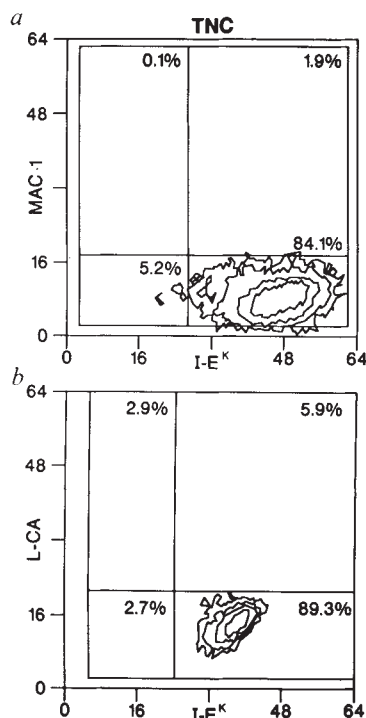


Fig. 4 Two-colour flow cytometric analysis of TNC surface marker expression. Each FACS profile is from TNC isolated from a total of 60 CBA/J mice: in *a*, 23,891 cells and in *b*, 8,619 cells were analysed using a Becton-Dickinson FACS 440. The TNC population was isolated as described in Fig. 2. This purified cell population was treated sequentially with anti-Thy 1.2 (AT83A, from F. Fitch), then guinea pig complement (Colorado Serum Co.) to remove any free thymocytes. The TNC were labelled with either rat-anti-MAC-1 (M1/70, from the American Type Culture Collection²⁵ or rat-anti-leukocyte common antigen (L-CA) (I3/2; ref. 26), plus biotinylated mouse-anti-I-E^{k/d} (14-4-4S; ref. 27). Secondary reagents used were fluorescein isothiocyanate (FITC)-conjugated goat anti-rat immunoglobulin cross-absorbed against mouse immunoglobulin (Cooper Biomedical) and streptavidin-phycoerythrin (PE)-conjugate (Chromoprobe). In these contour plots, PE and FITC fluorescence are plotted on the x and y axis respectively. Lines used to define positive versus negative cells were set against the outermost contour of controls where the primary antibodies were omitted. Both the T-ROS and M ϕ /DC population were positive for MAC-1 and L-CA (data not shown).

ized its cell-surface phenotype by two-colour fluorescence activated cell sorter (FACS) analysis. As shown in Fig. 4, the TNC population is class II-positive, but essentially negative for both leukocyte common antigen (a marker for bone marrow-derived cells) and MAC-1 (a marker for both thymic M ϕ and DC, but not for TNC)^{14,15}. These results indicate that the antigen presentation seen in our TNC preparation is not due to contamination from a bone marrow-derived APC. This raises the question of how the Hb reaches the epithelial cells. One possibility suggested by data reviewed in ref. 16 indicates that antigens can bypass the blood-thymus barrier and enter the cortical region by a transcapsular route.

Our experiments demonstrate that non-MHC self proteins are processed and presented by fetal thymic APCs at the time when an individual is acquiring tolerance to self. These findings have been extended to show that the APC responsible for this presentation of self molecules is not limited to one cell type, but instead can be accomplished by at least three distinct thymic cell populations. Two of these cell populations are traditional APCs, the M ϕ and the DC, but the third type, the cortical epithelial cell, has not been shown before to be able to present antigen. This novel finding forces us to revise our theories on the role of cortical epithelial cells in self tolerance and T-cell development^{17,18}. It must now be considered that for the self

antigen Hb, negative as well as positive selection may be occurring in the cortex, directed by non-bone marrow-derived cells. It will be interesting to see if tolerance to other self-proteins also follows this pattern.

We thank Kevin Roth for assistance with the photomicrographs, Gary Elliott and Bill Smith for flow cytometry, and Marian Simandl for preparation of the manuscript. This work was supported by the National Institutes of Health. P.M.A. is an Established Investigator of the American Heart Association.

Received 27 October; accepted 29 December 1988.

1. Billingham, R. E., Brent, L. & Medawar, P. B. *Nature* **172**, 603-606 (1953).
2. Matzinger, P., Zamoyska, R. & Waldmann, H. *Nature* **308**, 738-741 (1984).
3. Rammensee, H.-G. & Bevan, M. J. *Nature* **308**, 741-744 (1984).
4. Lorenz, R. G. & Allen, P. M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5220-5223 (1988).
5. Snodgrass, H. R., Kisielow, P., Kiefer, M., Steinmetz, M. & von Boehmer, H. *Nature* **313**, 592-595 (1985).
6. Sprent, J. & Webb, S. R. *Adv. Immun.* **41**, 39-133 (1987).
7. Wekerle, H. & Ketelsen, U.-P. *Nature* **283**, 402-404 (1980).
8. Wekerle, H., Ketelsen, U.-P. & Ernst, M. *J. exp. Med.* **151**, 925-944 (1980).
9. Kyewski, B. A. *Immun. Today* **7**, 374-379 (1986).
10. Kyewski, B. A., Rouse, R. V. & Kaplan, H. S. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5646-5650 (1982).
11. Beller, D. I. & Unanue, E. R. *J. Immun.* **118**, 1780-1787 (1977).
12. Beller, D. I. & Unanue, E. R. *J. Immun.* **124**, 1433-1440 (1980).
13. Kyewski, B. A., Fathman, C. G. & Kaplan, H. S. *Nature* **308**, 196-199 (1984).
14. van Vliet, E., Melis, M. & van Ewijk, W. *Cell. Immun.* **87**, 101-109 (1984).
15. Nabarra, B. & Papiernik, M. *Lab. Invest.* **58**, 524-531 (1988).
16. Nieuwenhuis, P. *et al. Immun. Today* **9**, 372-375 (1988).
17. Janeway, C. A. *Nature* **335**, 208-210 (1988).
18. Marrack, P. & Kappler, J. *Immun. Today* **9**, 308-315 (1988).
19. Allen, P. M. & Unanue, E. R. *J. Immun.* **132**, 1077-1079 (1984).
20. Lorenz, R. G. & Allen, P. M. *Immun. Rev.* **106**, 115-127 (1988).
21. Jenkinson, E. J., van Ewijk, W. & Owen, J. J. T. *J. exp. Med.* **153**, 280-292 (1981).
22. Wong, P. M. C., Chung, S.-W., Reicheld, S. M. & Chui, D. H. K. *Blood* **67**, 716-721 (1986).
23. Kyewski, B. A. & Kaplan, H. S. *J. Immun.* **128**, 2287-2294 (1982).
24. Kyewski, B. A., Fathman, C. G. & Rouse, R. V. *J. exp. Med.* **163**, 231-246 (1986).
25. Springer, T., Galfre, G., Secher, D. S. & Milstein, C. *Eur. J. Immun.* **9**, 301-306 (1979).
26. Trowbridge, I. S. *J. exp. Med.* **148**, 313-323 (1978).
27. Ozato, K., Mayer, N. & Sachs, D. H. *J. Immun.* **124**, 533-540 (1980).

Clonal deletion of B lymphocytes in a transgenic mouse bearing anti-MHC class I antibody genes

David A. Nemazee* & Kurt Bürki†

* Basel Institute for Immunology, CH-4005 Basel, Switzerland

† Preclinical Research, Sandoz Ltd, CH-4002 Basel, Switzerland

B lymphocytes can be rendered specifically unresponsive to antigen by experimental manipulation *in vivo* and *in vitro*¹⁻⁶, but it remains unclear whether or not natural tolerance involves B-cell tolerance because B cells are controlled by T lymphocytes, and in their absence respond poorly to antigen (reviewed in ref. 7). In addition, autoantibody-producing cells can be found in normal mice and their formation is enhanced by B-cell mitogens such as lipopolysaccharides⁸⁻¹². We have studied B-cell tolerance in transgenic mice using genes for IgM anti-H-2^k MHC class I antibody. In H-2^d transgenic mice about 25-50% of the splenic B cells bear membrane immunoglobulin of this specificity, and abundant serum IgM encoded by the transgenes is produced. In contrast, H-2^k × H-2^d (H-2^{d/k}) transgenic mice lack B cells bearing the anti-H-2^k idotype and contain no detectable serum anti-H-2^k antibody, suggesting that very large numbers of autospecific B cells can be controlled by clonal deletion.

We cloned genes for the transgenic (Tg) anti-H-2K^d antibody from the DNA of the BALB/c IgG-producing hybridoma 3-83 (ref. 13), and constructed a chimaeric heavy-chain gene using a mouse C μ constant-region DNA clone (see Fig. 1*a*). The founder male Tg mouse was derived from an embryo that was homozygous for the H-2^d haplotype (H-2^{d/d}). Its serum had a high anti-H-2^k cytotoxic antibody titre and contained at least 50 μ g ml⁻¹ of the 3-83 idotype (id), as recognized by the rat monoclonal anti-idiotypic antibodies 35B and 54.1 that bind

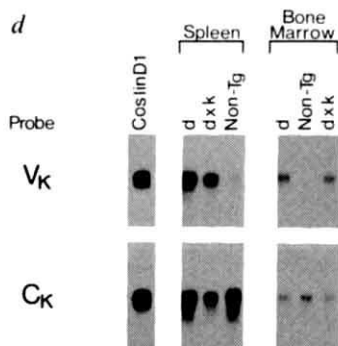
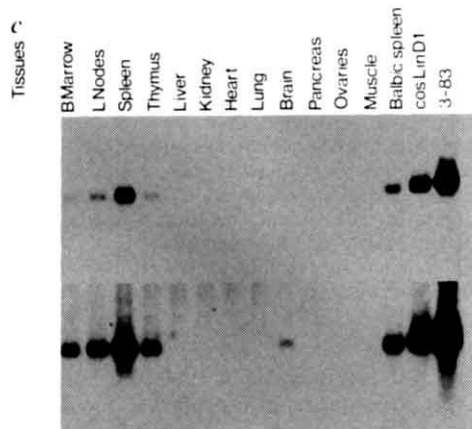
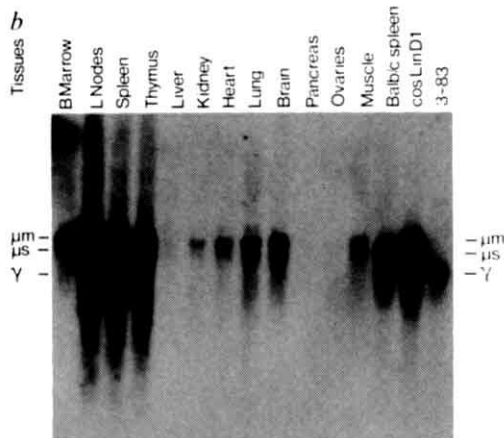
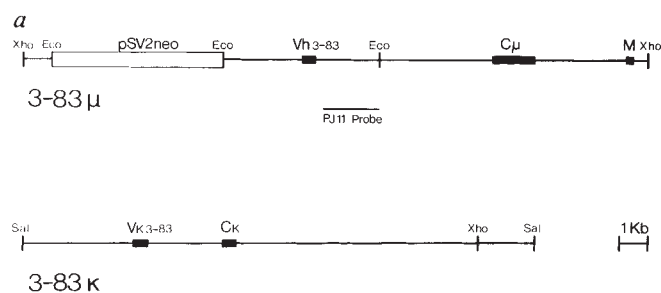


Fig. 1 Production and RNA analysis of transgenic mice. **a**, Genes used for generation of transgenic mice. Open box indicates the pSV2neo vector sequences²¹, and the pJ11 (ref. 22) probe is indicated. **b**, **c** And **d**, Northern analysis of H-2^d Tg tissues using 3-83 V μ , V κ and C κ probes: **b**, V μ probe. 10 μ g RNA (3-83 RNA, 1 μ g) was loaded in each lane. Cos Lin D1 was a myeloma transfectant expressing chimaeric 3-83 genes; **c**, V κ probe. As in **b**, except that only 5 μ g lymph node RNA was run. The two panels show 1- and 7-day exposures of the same blot; **d**, 5 μ g RNA per lane. The filter was first hybridized with V κ probe (top), exposed, stripped and rehybridized with C κ probe.

Methods. The VDJ μ segment was isolated from λ gt-wes *Eco*RI library using J-C μ intron-specific probe pJ11. The intact V κ 3-83 gene was isolated from an EMBL3 *Sau*3A partial digest library. The VDJ μ 3-83 was ligated into plasmid pR μ R1 containing the genomic C μ gene. To prepare DNA for microinjection, 3-83 μ plasmid was linearized at the unique *Xho*I site. The 3-83 κ DNA was excised with *Sal*I, which cuts in polylinker regions flanking the insert, and the insert was isolated by electrophoresis through low-melting-temperature agarose. Transgenic mice were produced as described^{23,24} from DBA/2 $\times$ B10D2F₂ zygotes: about 2 pl DNA solution containing about 500 molecules²⁵ were microinjected. Purified RNAs were electrophoresed through agarose-formaldehyde gels and transferred to nitrocellulose²⁶. Hybridization was at 42 °C in 50% formamide, 1 M NaCl, 10 $\times$ Denhardt's solution, 50 mM Tris-HCl, pH 7.5, 0.1% Na₂P₂O₇, 1% SDS, 150 μ g ml⁻¹ salmon sperm DNA. The V μ probe was a 400-base pair *Eco*RV-*Bst*EII fragment that included most of the coding sequence. The V κ probe was a 250-base pair *Hind*III-*Bam*HI fragment encompassing the 5' portion of the gene. The C κ probe was a *Hpa*I-*Bgl*II fragment that included most of the coding sequence. The final wash was for 30 min in 0.3 $\times$ SSC, 0.1% SDS at 65 °C.

an epitope formed by the combination of the 3-83 heavy and light chains¹⁴.

Mice of this transgenic line (designated Tol 1) have 8-10 copies of the microinjected genes, which are apparently integrated at a single locus in an autosome. The 3-83 μ messenger RNA was found at high concentrations in lymphoid organs and present in low amounts in several other tissues (Fig. 1b), whereas the 3-83 kappa mRNA expression was limited to lymphoid tissues, except in the brain, where it was present at a low level (Fig. 1c). The transgenic μ message is expressed in non-lymphoid cells and at very high amounts in the thymus. It is likely that T cells express large amounts of this message. Two-colour immunofluorescence analysis showed that in H-2-d/d Tg spleen the 3-83 idiotype is found only on B cells (Fig. 2a), indicating a tissue-specific expression of the 3-83 IgM protein.

To determine whether the presence of the H-2^k class I antigens affects 3-83 protein expression, we analysed three litters of [(BALB/c (H-2^d) $\times$ BALB/k (H-2^k)] $\times$ Tol 1 (H-2^d) mice (Table 1). The segregation data showing the expected mendelian ratios, and the large litter sizes (8, 9 and 9 offspring per (BALB/c $\times$ BALB/k) F₁ mother respectively), all suggest that no death occurred among H-2-d/k embryos bearing the transgenes. H-2-d/d Tg sera contained high serum concentrations of 3-83 id, which was not detectable in H-2-d/k Tg or in any of the non-Tg littermates, levels being at least 200-fold less (Table 1). 3-83 id made up 61.6% of the total IgM in H-2^d Tg mice. In H-2-d/k Tg mice the IgM levels in the sera were reduced to only 26% of that of the H-2^d Tg mice.

To determine the basis for the apparent tolerance in the H-2-d/k Tg mice, we analysed non-Tg, H-2-d/d Tg and H-2-d/k Tg spleen and bone-marrow cells by cell-surface immunofluorescence (Fig. 2). A quarter to a half of the splenic B cells in H-2-d/d Tg bear the 3-83 id, whereas id-positive cells were undetectable in the H-2-d/k Tg mice; the B cells with the pre-determined transgene-encoded anti-self specificity were absent. Non-Tg mice were completely id-negative. We found 3-83 id-positive cells in the spleen, lymph nodes and bone marrow, but not in the thymus of H-2-d/d Tg mice, whereas H-2-d/k Tg mice lacked idiotype-bearing cells in all these tissues. Given the background levels of staining we can be confident that the

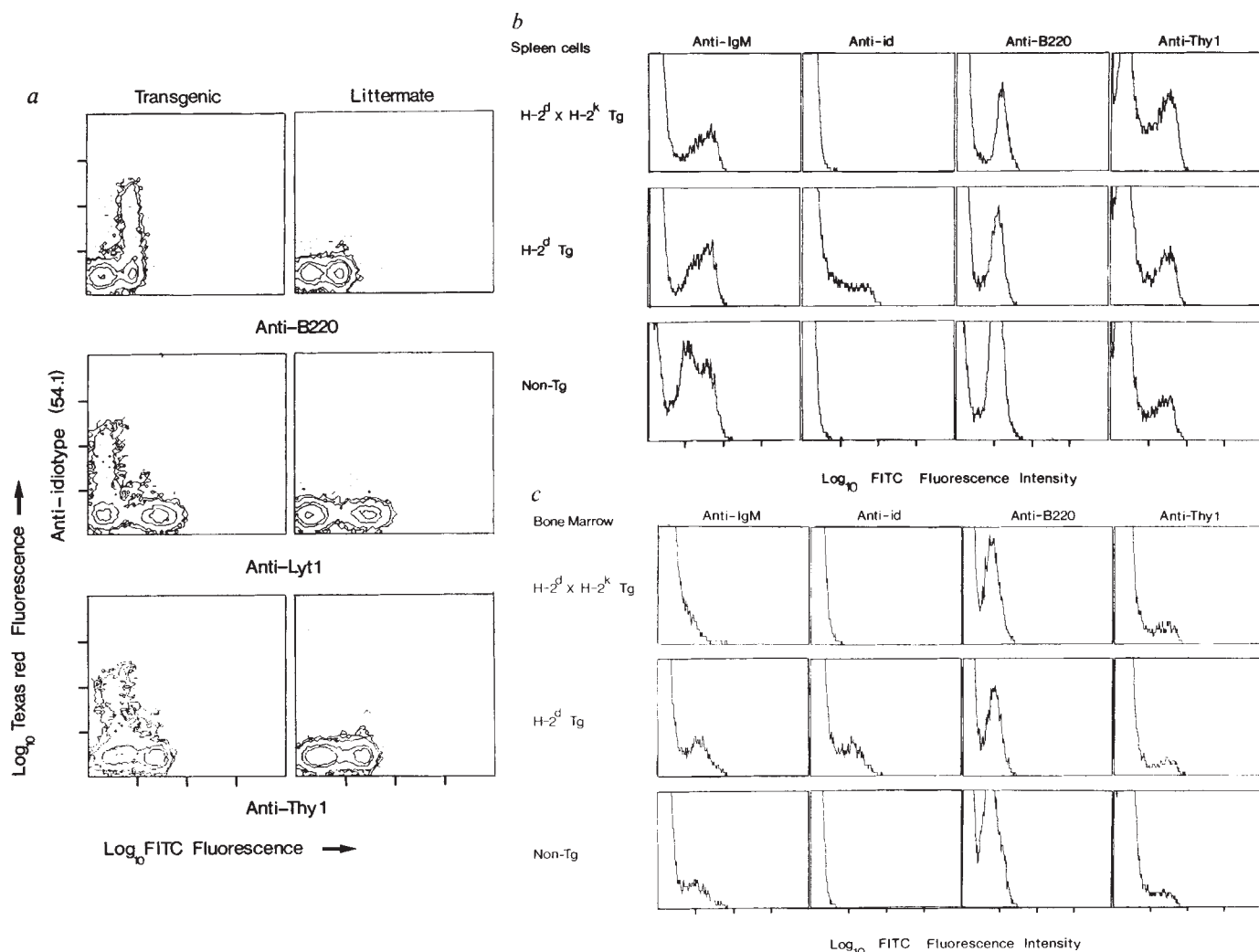


Fig. 2 Flow cytometry analysis. *a*, Two-colour staining of H-2-d/d Tg and control littermate spleen cells. Cells were first incubated with a mixture of biotin-54.1 and fluorescein isothiocyanate (FITC)-coupled antibodies washed and incubated with Texas red-avidin. Samples were analysed using a Becton-Dickinson fluorescence-activated cell-sorting FACS 440 machine. *b* and *c*, Cells were stained with monoclonal rat anti-IgM (M41) or anti-idiotypic (54.1) followed by FITC-labelled mouse anti-rat IgG, or with directly coupled FITC-anti-B220 (RA3-3A1) or FITC-anti-Thy1. Negative control staining with BAG1 (a rat IgG1 monoclonal) gave a result virtually identical to that achieved with anti-idiotypic staining of non-transgenic cells. Samples were analysed on a Becton-Dickinson FACSCAN machine.

Methods. All samples were depleted of erythrocytes by treatment in Gey's solution. Dilutant and wash solution in all cases was buffer A (see legend to Table 1). Analysis in *b* and *c* was in buffer A supplemented with 5 $\mu\text{g ml}^{-1}$ propidium iodide to help exclude dead cells from analysis.

frequency of 3-83 id-bearing cells in the H-2-d/k Tg spleens was at least 50-fold lower than in H-2-d/d spleens.

It is unlikely that the capping of cell-surface idiotype bound to H-2^k in H-2-d/k Tg mice would explain the lack of id-positive cells in the spleens of these mice, because in all cases the percentage of IgM-bearing cells closely matched the percentage of cells staining with an independent B-cell marker (B220) and the density of IgM on H-2-d/k and H-2-d/d Tg B cells was comparable (Fig. 2*b*). By contrast, the staining pattern seen with H-2-d/k Tg bone marrow cells (Fig. 2*c*) with its reduced density of IgM and low % IgM:B220 ratio indicates that modulation of cell surface IgM could have occurred, probably after binding to H-2^k class I molecules. We note in this regard that surface immunoglobulin is rapidly capped and then internalized, or is shed within minutes after cross-linking, and that bone-marrow B cells in particular fail to re-express surface immunoglobulin. This has been proposed as a model system for tolerance¹⁵.

To confirm that the lack of 3-83 id in the sera of H-2-d/k Tg mice was due to tolerance rather than to adsorption to H-2^k class I molecules, we analysed frozen sections of spleens for cytoplasmic expression of 3-83 id and IgM using two-colour immunofluorescence. A large fraction of the IgM-producing

plasma cells in the H-2-d/d mice produce 3-83 id, whereas almost none of those in H-2-d/k mice were idiotype-positive (Fig. 3). These experiments also confirmed the lack of surface idiotype-positive cells in H-2-d/k Tg and the presence of such cells as a subset of IgM-bearing cells in the H-2-d/d Tg mice. Taken together, these results indicate that 3-83 id-producing cells in the H-2-d/k mice were largely eliminated or excluded from the spleen. The deletion probably occurred at some time shortly after they first expressed their immunoglobulin receptors, so most of them never developed into antibody-secreting cells. This would support reports demonstrating functional inactivation of B cells in tolerant animals^{2,3,5}.

Despite the absence of 3-83 id-protein expression, H-2^k Tg spleen continues to produce Tg light chain (Fig. 1*d*) and heavy chain mRNA, although levels are reduced compared with H-2^d Tg. As idiotype absorption is ruled out, it seems that the RNA is produced either by non-B cells or as a result of B cells expressing the transgenes which have either mutated these genes or are producing functional immunoglobulin protein chains in addition; these would be encoded by endogenous genes, giving rise to idiotype-negative molecules when paired with Tg chains. Another possibility is that a small, but highly metabolically

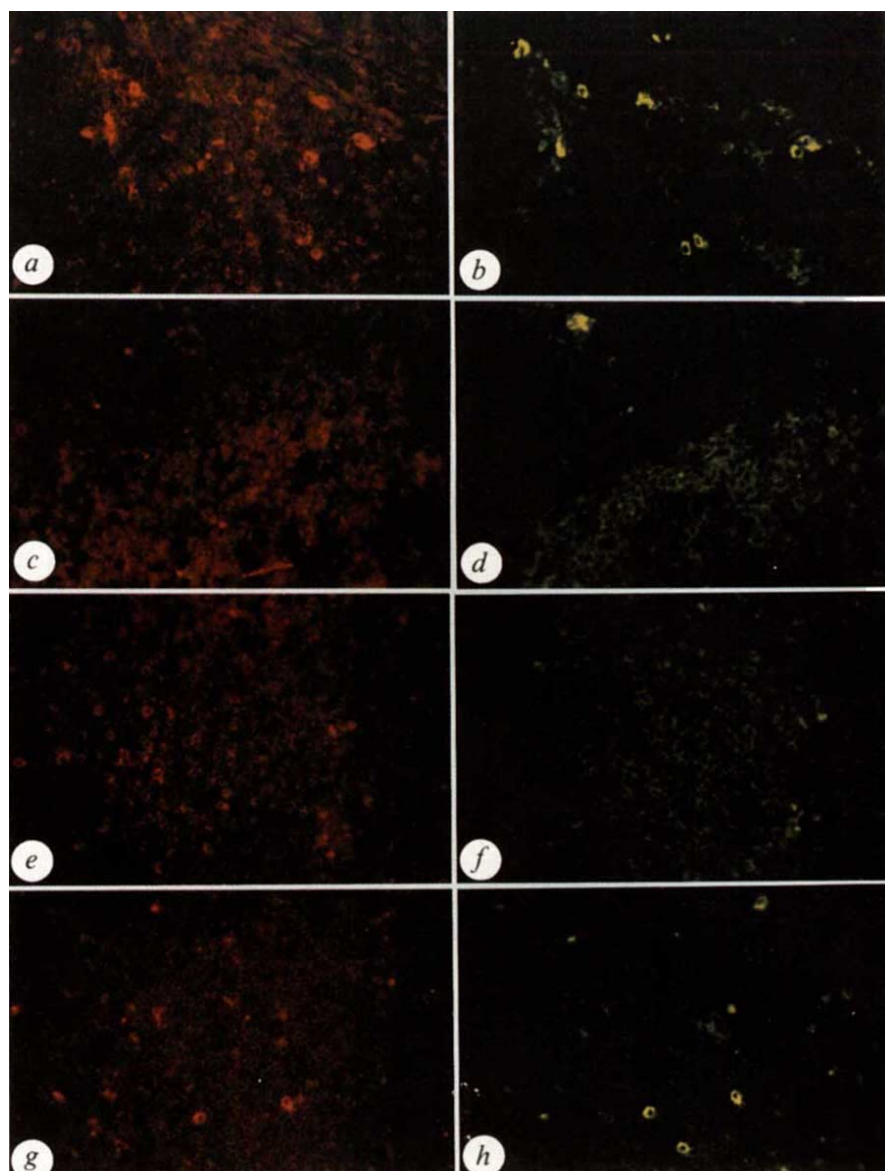


Fig. 3 Two-colour immunofluorescence. *a*, H-2-d/d spleen stained with anti-idiotypic; *b*, same section as in *a*, stained with anti- μ ; *c*, H-2-d/k spleen anti-idiotypic; *d*, same section as in *c*, anti- μ ; *e*, field containing a B-cell area from H-2-d/d spleen, anti-idiotypic; *f*, same section as in *e*, anti- μ ; *g*, H-2-d/d thymus, anti-id; *h*, same as *g*, anti- μ .

Methods. Frozen sections from H-2-d/d spleen (*a*, *b*, *e* and *f*); H-2-d/k spleen (*c*, *d*); H-2-d/d thymus (*g*, *h*) were incubated for 30 min at 4° with a mixture of FITC-rat-anti-IgM and biotin-54.1 in PBS, washed in PBS and incubated with Texas-red avidin for 30 min at 4°. Photographs were taken through a Leitz fluorescence microscope at magnification 360 $\times$ using the appropriate filters to detect Texas red (*a*, *c*, *e*, *g*) or FITC (*b*, *d*, *f*, *h*) fluorescence.

active, population of transgene-expressing B cells persists.

In all H-2-d/d Tg animals tested so far, a large fraction of the splenic B cells (50–75%) were idiotype-negative. The fact that not all B cells in the H-2-d/d Tg mice bear the idiotype is to be expected on the basis of earlier studies of allelic exclusion in immunoglobulin-transgenic mice (reviewed in ref. 16). Interestingly, almost all H-2-d/d Tg bone-marrow IgM⁺ B cells were also idiotype-positive, indicating that rare variant B cells may be strongly selected in peripheral lymphoid organs. The idiotype-negative B cells in Tol 1 mice could be the result of somatic mutation of transgenes, or of the association of the Tg immunoglobulin heavy or light chains with light or heavy chains encoded by endogenous genes.

Our conclusion that deletion is taking place in H-2-d/k Tg mice rests on the following data. The peripheral B cells in these mice (as defined by the marker B220) all bear high concentrations of surface IgM but no detectable idiotype, and their frequency is reduced by ~50% compared with H-2-d/d Tg control littermates. Furthermore, in the bone marrow there is a striking difference in surface IgM levels between H-2-d/d Tg and H-2-d/k Tg mice, indicating that capping or down-regulation of immunoglobulin is occurring in the latter. If idiotype-producing B cells populate the spleen in H-2-d/k Tg mice, the staining patterns of the bone marrow and spleen cells of these mice would be expected to be similar. These results show that

there must be a significant reduction in the number of id-producing B cells in the spleen and lymph nodes of H-2-d/k Tg mice, but they cannot exclude the possibility that a small number of such cells are spared. If residual id-positive B cells exist, they could well be in an activated state (having presumably encountered antigen) and might contribute to the levels of transgene-encoded mRNA observed in H-2-d/k spleen.

The expression of H-2D^k and K^k on the B cells of H-2-d/k Tg mice is probably not required for clonal deletion. This conclusion is based on radiated bone-marrow chimaeras of the type H-2^dTg $\rightarrow$ H-2^d $\times$ H-2^k, indicating that the same bone marrow inoculum can give rise to id-positive B cells in an H-2^d environment, but not in an H-2^d $\times$ H-2^k environment (data not shown). It is possible that in some cases the interaction of antigen receptor and antigen in the cell (perhaps in the endoplasmic reticulum) could permit specific signal transduction leading to tolerance; this might be an important mechanism for eliminating a wide variety of anti-self specificities and even play a part in T-cell tolerance.

In contrast to our results, a recent report¹⁷ claims that functional inactivation of autoreactive B cells occurs without physical deletion in mice transgenic for anti-hen egg lysozyme. As these authors point out, the lack of clonal deletion in their system may be the result of the low concentration or univalent form of the autoantigen (transgene-encoded hen-egg lysozyme)

Table 1 Serum concentrations of immunoglobulins in (BALB/c × BALB/k) × Tol 1 mice

	IgM μg ml ⁻¹	3-83 Idiotypic μg ml ⁻¹	Number of animals tested
Non-transgenic	846 ± 301*	<0.3 (<0.035%)†	13
H-2 d/d transgenic	151 ± 42	93 ± 34 (61.6%)	7
H-2 d/k transgenic	40 ± 21	<0.3 (<0.75%)	6

Sera were taken 22 days after birth. IgG titres were uniformly high in all mice and may have been of maternal origin. Radioimmunoassay was performed as follows: polyvinylchloride wells were coated overnight at 4 °C with 50 μl per well of PBS containing 5 μg ml⁻¹ purified 54.1 (rat anti-idiotypic), m41 (rat anti-μ) or YA2 (rat anti-IgG). Wells were washed 3 times and then incubated for 30 min with PBS, 1% bovine serum albumin, 0.02% sodium azide (buffer A). Various concentrations of serum or control antibodies diluted in buffer A were applied to the wells and incubated for 3 hours at 4 °C. After washing, 2.5 ng ¹²⁵I-labelled 33-18-12 (rat anti-mouse kappa) diluted in buffer A were applied to the wells for two hours at 4 °C. The wells were extensively washed with PBS and then with tap water and the bound radioactivity was measured in a gamma counter. As a positive control and to calibrate the IgM and idiotype assays, Cos lin D1 monoclonal antibody was used. This antibody is derived from the supernatant of Sp2/0 myeloma cells transfected with a cosmid containing the DNA fragments shown in Fig. 1 (ref. 14). For segregation analysis of transgenes and MHC genes, data were obtained from Southern blots using BamHI-digested tail DNA. A number of hybridization probes, including pJ11 (see Fig. 1), could unambiguously distinguish Tg from non-Tg individuals. The complementary DNA probe p2IIa (ref. 20), which hybridizes to many class I genes, was used to distinguish H-2-d/k and H-2-d/d mice on the basis of three different H-2^k-specific fragment length polymorphisms.

* Errors listed are ± 1 standard deviation.

† Per cent IgM that is of the 3-83 idiotype.

used. It may be that these B cells do not encounter the physiologically relevant form or concentration of hen-egg lysozyme until they are relatively mature, and hence must be dealt with in a different way, for example by suppression. It should also be pointed out that in the presence of antigen the numbers of hen-egg lysozyme-specific B cells in the transgenic mice are lower than in the controls lacking antigen, suggesting that either there is deletion of part of the clone (perhaps a particular B-cell subset) or that the lifespan of the entire clone is reduced. The MHC class I antigens used in our system are cell-surface proteins that could probably permit extensive cross-linking of immunoglobulin receptors on reactive B cells arising in the bone marrow. A signal at this developmental stage might lead directly to B-cell inactivation and death.

The *in vivo* deletion of self-reactive B cells agrees with recent experiments indicating that T cells bearing receptors for self-MHC antigens are also physically deleted¹⁸. Thus our results would support the hypothesis that B cells are made tolerant by antigen in the absence of antigen-reactive T cells¹⁹. Our Tg mice should allow us to explore more carefully the cellular and biochemical events leading to tolerance in the B-cell compartment.

The authors thank Drs Fritz Melchers, Martin Flajnik, Chris Coleclough and Howard Etlinger for critical reading of the manuscript, Drs Rolink and L. Forni for antibody reagents and C. Guet and U. Steinmann for technical assistance. We thank David Gray for help with section-staining and Janette Millar for preparation of the manuscript. The Basel Institute for Immunology was founded and is supported by F. Hoffman-La Roche & Co. Ltd, Basel, Switzerland.

Received 12 September; accepted 22 December 1988.

1. Dresser, D. W. *Immunology* **5**, 378-388 (1962).
2. Chiller, J. M., Habicht, G. S. & Weigel, W. O. *Science* **171**, 813-815 (1970).
3. Borel, Y. *Nature new Biol.* **230**, 180 (1971).
4. Etlinger, H. M. & Chiller, J. M. *Cell Immun.* **33**, 297-308 (1977).
5. Teale, J. M. & Klinman, N. R. *Nature* **288**, 385-387 (1980).

6. Pike, B. L., Kay, T. W. & Nossal, G. J. V. *J. exp. Med.* **152**, 1407-1412 (1980).
7. Nossal, G. J. V. *Proc. R. Soc.* **288**, 225-240 (1986).
8. Moller, G. *et al. J. exp. Med.* **143**, 1429-1438 (1982).
9. Harris, D. E., Cairns, L., Rosen, F. S. & Borel, Y. *J. exp. Med.* **156**, 567-584 (1982).
10. Jemmerson, R. R. W., Morrow, P. R., Klinman, N. R. & Patterson, Y. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1508-1512 (1985).
11. Karay, S., Lymberi, P., Avranmeas, S. & Continho, A. *Scand. J. Immun.* **23**, 475-480 (1986).
12. Rolink, A. G., Radaszkiewicz, T. & Melchers, F. *J. exp. Med.* **165**, 1675-1687 (1987).
13. Ozato, K., Mayer, N. & Sachs, D. H. *J. Immun.* **124**, 533-540 (1980).
14. Nemazee, D. A. in *Proc. EMBO Workshop on Tolerance* (eds Matzinger, P. *et al.*) Vol. 2, 52-54 (Editiones Roche, Basle, 1987).
15. Sidman, C. & Unanue, E. *Nature* **257**, 149-151 (1975).
16. Storb, U. A. *Rev. Immun.* **5**, 151-174 (1987).
17. Goodnow, C. C. *et al. Nature* **334**, 676-682 (1988).
18. Kappler, J. W., Roehm, N. & Marrack, P. *Cell* **49**, 273-280 (1987).
19. Bretscher, P. & Cohn, M. *Science* **169**, 1042-1049 (1970).
20. Steinmetz, M. *et al. Cell* **24**, 125-134 (1981).
21. Southern, P. J. & Berg, P. *J. molec. appl. Genet.* **1**, 327-341 (1982).
22. Marcu, K. B. *et al. Cell* **22**, 187-196 (1980).
23. Gordon, J. W. & Ruddle, F. W. *Meth. Enzym.* **101**, 411-433 (1983).
24. Bürki, K. *Mon. Devl Biol.* **19**, 1-73 (1986).
25. Brinster, R. L., Chen, H. Y., Trumbauer, M. E., Yagle, M. K. & Palmiter, R. D. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4438-4442 (1985).
26. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: a Laboratory Manual* 196-203 (Cold Spring Harbor Laboratory, New York, 1982).

Spatial and temporal expression of the retinoic acid receptor in the regenerating amphibian limb

Vincent Giguère*†, Estelita S. Ong*, Ronald M. Evans* & Clifford J. Tabin‡

* Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

‡ Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

Retinoic acid is known to have dramatic effects on vertebrate limb pattern in development and regeneration, supporting a model in which a gradient of retinoic acid serves as a morphogen to differentially supply positional information to a developing limb. The discovery of a retinoic acid receptor (RAR)^{1,2} and its homology to the steroid and thyroid hormone receptors³ provided a potential molecular mechanism for limb morphogenesis⁴. One prediction of this model is that the receptor must be expressed in the developing and regenerating limb anlage. We investigated the expression of the RAR in the adult newt, *Notophthalmus viridescens*, whose amputated limbs are capable of regenerating and upon which retinoic acid can act to alter pattern. We report the cloning of cDNAs encoding a functional newt RAR and the localization of high and uniform levels of RAR mRNA specifically in the regenerating cells that control limb pattern. These results indicate that the morphogenic field is established through differential activation of pre-existing retinoic acid receptors rather than differential expression of the RAR gene.

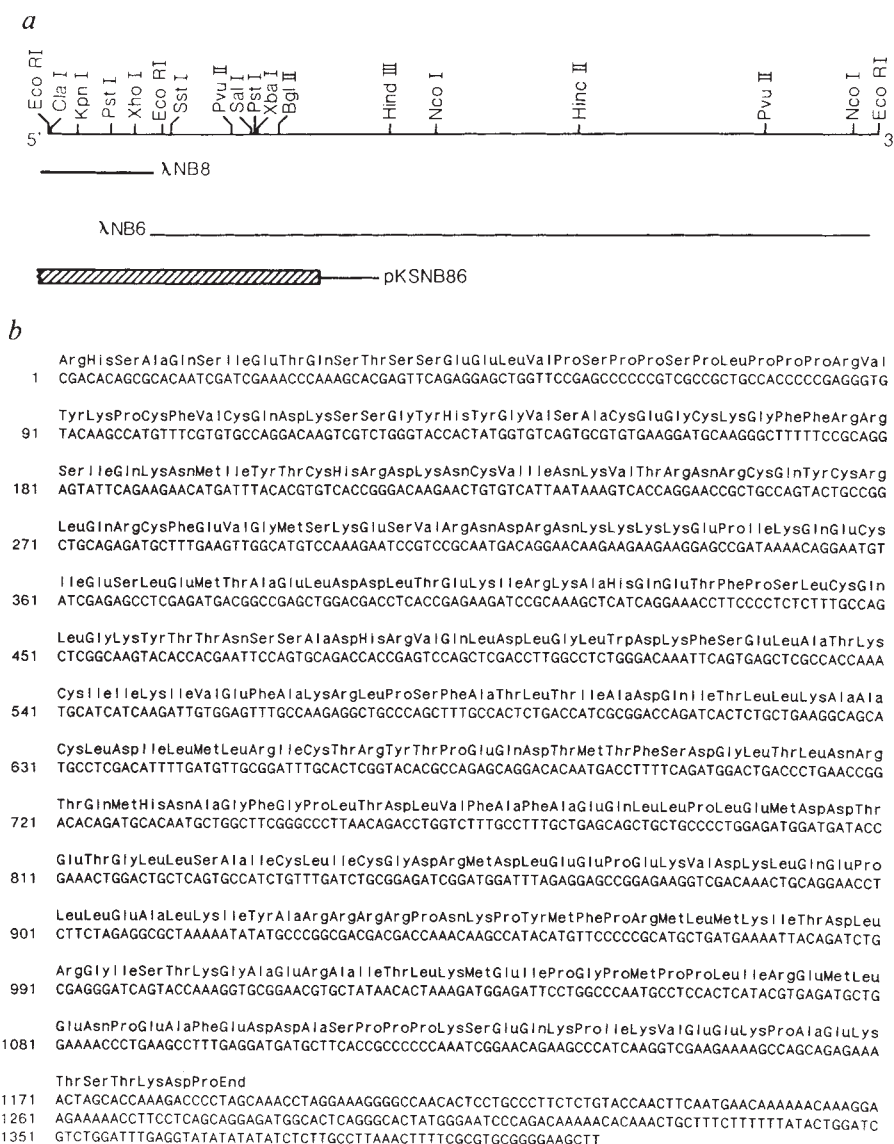
To confirm the existence of a RAR in the regenerating newt limb, a λgt10 cDNA library made from early bud stage, proximal hindlimb blastema poly(A)⁺ RNA was screened with a human RARα cDNA¹ as probe. Two positive clones, designated λNB6 and λNB8, were isolated and their *Eco*R1 inserts subcloned in a pGEM4 plasmid (Fig. 1a). The composite nucleotide and predicted amino-acid sequences of this putative newt RAR cDNA are shown in Fig. 1b. In humans, two distinct RAR gene products hRARα^{1,2} and hRARβ^{5,6} have been identified. The newt protein is highly related to both human receptors with greater apparent similarity to the β-form, especially in the N-terminal region.

To test whether the newt RAR is a functional retinoic acid receptor, the expression plasmid RSnRAR was cotransfected in

† Present address: Division of Endocrinology, Research Institute, The Hospital for Sick Children, 555 University Avenue, Toronto, Canada M5G 1X8

Fig. 1 a, Schematic representation of subcloned *EcoRI* fragments derived from newt RAR cDNA clones λ NB6 and λ NB8. The composite cDNA pKSNB86 is represented with coding (stippled portion) and non-coding (thin line) sequence indicated. Common six-nucleotide restriction enzyme sites are drawn above the linear map. **b**, DNA sequence and predicted amino-acid sequence of the newt RAR cDNAs. Nucleotide sequence of the composite cDNA up to the *HindIII* site with deduced amino acids given above the long open reading frame.

Methods. To isolate newt RAR cDNA clones, the entire insert of clone λ HK1R encoding the human RAR α (ref. 1) was used as a hybridization probe to screen a newt blastema λ gt10 cDNA library. The hybridization mixture contained 35% formamide, $1 \times$ Denhardt's $5 \times$ SSPE (0.15 M NaCl, 0.01 M NaH₂PO₄, 0.001 M EDTA), 0.1% SDS, $100 \mu\text{g ml}^{-1}$ denatured salmon sperm DNA and 10^6 c.p.m. ml^{-1} of ^{32}P -labelled DNA probe. Duplicate nitrocellulose filters were hybridized at 42°C for 16 hours, washed three times for 20 minutes each in $2 \times$ SSC, 0.1% SDS at 55°C and autoradiographed at -70°C with an intensifying screen. The *EcoRI* insert of two cDNA clones, λ NB6 (3,050 base pairs) and λ NB8 (481 base pairs), were subcloned in pGEM4 and analysed further. The *EcoRI* inserts were digested with a number of restriction enzymes and the resulting fragments subcloned into the M13 sequencing vectors mp18 and mp19 and sequenced by the chain termination techniques¹⁶. The composite cDNA pKSNB86 was generated as follows: the *EcoRI* insert of λ NB8 was first introduced in the *EcoRI* site of vector pBluescriptKS (Stratagene) to generate pKSNB8. Then, a 930-base pair *EcoRI*/*HindIII* fragment from λ NB6 was introduced in pKSNB8 to yield pKSNB86. Transfection of CV-1 cells, treatments with retinoic acid, luciferase and β -galactosidase assays were performed as described¹⁷.



the mammalian cell line CV-1 with the reporter plasmid Δ MTV-TREp-LUC that has been shown to be responsive to human thyroid hormone and retinoic acid receptors⁷ (Fig. 2a). The results in Fig. 2b show that luciferase activity is induced in a dose-dependent way, as expected for a functional receptor. The ED₅₀ (50% effective dose) for retinoic acid action in this transcriptional assay is ~ 4 nM which is well within the range of an ED₅₀ for the various biological effects of retinoic acid observed *in vitro*⁸.

In axolotyl regeneration the primary effect of exogenously applied retinoic acid is to cause proximal-distal duplications⁹. To examine this parameter in the newt, 30 μl of 10 mM retinoic acid was injected intraperitoneally at sequential stages during regeneration (see Table 1). The limbs were scored without dissection, and only duplication of the lower and upper arm was scored. Thus, the percentages of duplication shown are probably underestimates of retinoic acid action as carpal duplications were not assayed. Applied retinoic acid is most effective in

Table 1 Proximal-distal duplications caused by retinoic acid treatment

Morphological stage at time of injection	Day of injection	Survival	No arm duplication	Lower arm duplication	Lower + upper arm duplication	% Duplication
Control	—	10/10	20	0	0	0
Wound healing	1	4/10	6	1	1	25
Early dedifferentiation	4	5/10	2	5	3	80
Late dedifferentiation	7	5/10	0	6	4	100
Later dedifferentiation	10	4/10	2	1	5	75
Growth-early bud	13	5/10	4	4	2	60
Growth-medium bud	16	6/10	7	2	3	42
Redifferentiation/morphogenesis-late bud	19	2/10	3	0	1	25
Redifferentiation/morphogenesis-palette	22	5/10	9	1	0	10
Redifferentiation/morphogenesis-early digit	25	4/10	8	0	0	0

RNA (30 μl of 10 mM RA) was injected intraperitoneally at different times after bilateral amputation.

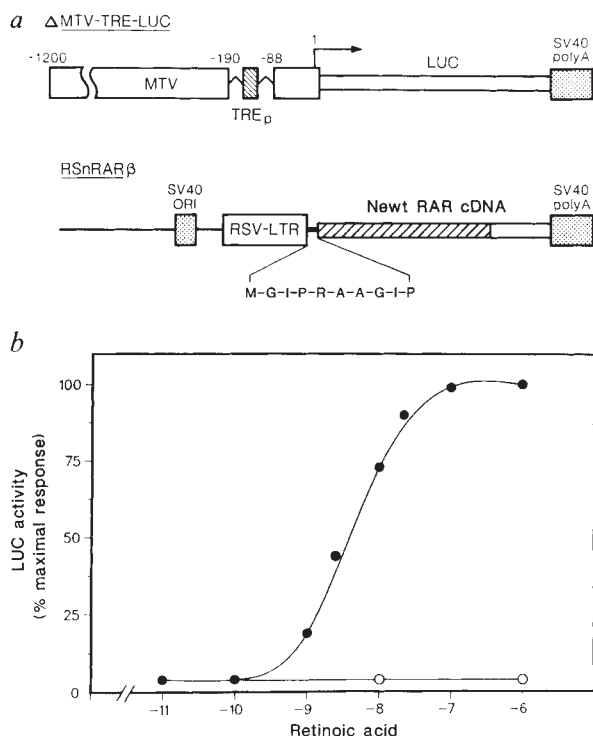


Fig. 2 Transcriptional activation by the newt RAR. **a**, Schematic representation of the retinoic acid-responsive reporter gene and expression vector used in this experiment. The reporter plasmid Δ MTV-TREp-LUC consists of the mouse mammary tumour virus long terminal repeat (MTV-LTR) from which glucocorticoid-responsive elements were deleted and replaced by a synthetic 33-base pair oligonucleotide encoding a palindromic variant of the rat growth hormone thyroid-hormone responsive element¹⁸, the luciferase gene (LUC) and the SV40 polyadenylation signal. Numbering is relative to the start site of transcription. The expression vector pRSnRAR contains the newt RAR cDNA driven by the Rous sarcoma virus LTR. The coding sequence of the newt RAR cDNA is shown as a stippled box. Additional amino acids encoded in the synthetic linker are shown. **b**, Induction of luciferase activity by the newt RAR in response to increasing doses of retinoic acid. Luciferase activity was measured using CV-1 cells extracts obtained by cotransfection of the reporter plasmid Δ MTV-TREp-LUC, the expression vector carrying the newt RAR cDNA in sense (●) or antisense (○) orientation and the plasmid RSV- β GAL as internal control. The levels of luciferase activity were plotted as percentages of the maximal response observed in this particular experiment. No luciferase activity was detected in absence of retinoic acid.

Methods. The expression vector pRSnRAR was constructed as follows. The parent expression vector pRShGR (ref. 17) was modified to include a *Nco*I site 5' of the hGR coding sequence by insertion of a *Nco*I linker (5'CCCATGGG3') at the *Kpn*I site to generate pRShGRN. The introduction of a *Nco*I site provided an AUG initiator codon. A similar *Nco*I site was introduced at the *Bam*HI site of pKSNB86 5' of the newt RAR coding sequence which gave pKSNB86N. The DNA sequence encoding hGR was then excised from pRShGRN by digestion with *Bam*HI, followed by end-filling to generate a blunt end, and then by *Nco*I. The sequence encoding the newt RAR was excised from pKSNB86N by digestion by *Cla*I, followed by end-filling, and *Nco*I, and inserted into *Nco*I/*Bam*HI-digested pRShGRN to generate pRSnRAR. Reporter plasmid Δ MTV-TREp-LUC was constructed by replacing the CAT gene present in the construct Δ MTV-TREp-CAT⁷ with the luciferase gene obtained from pSVOA/L- Δ 5' (ref. 19).

altering normal regeneration during a short period of time peaking on day 7 (during dedifferentiation) and therefore, endogenous retinoic acid is probably most influential in pattern formation at this time.

On the basis of this data, *in situ* hybridization was used to determine the expression of the RAR 7 days after amputation. Limb blastemas were cryostat-sectioned along a plane parallel to their dorsal surface and hybridized to either sense or antisense RNA. The region of the clone from which the probe was transcribed encodes the retinoic acid-binding domain of the RAR which, unlike the conserved DNA-binding domain, does not cross-react with other members of the thyroid/steroid receptor superfamily. After hybridization, the sections were exposed for 10 days. No specific hybridization was observed with the sense probe (data not shown). Examples of antisense hybridization are shown in Fig. 3. Mesenchymal cells directly under the wound epidermis express the RAR to a greater extent than the normal limb below the regenerating region. It is the mesenchymal cells rather than the epidermal cells that are the targets of retinoic acid action¹⁰. Thus, RAR mRNA expression is selectively activated, in the correct cell type and at the correct point during regeneration for it to mediate the effects of retinoic acid as a morphogen.

We also wanted to know if the receptor itself is expressed in a spatial gradient that corresponds to either the proximal/distal axis or the anterior/posterior axis which retinoic acid also affects¹¹. To look for potential anterior/posterior gradients 64 blastemas from 38 different animals were scored. No reproducible gradient of RAR expression could be associated with the anterior/posterior axis.

To score for graded expression along the proximal/distal axis we compared sections from blastemas resulting from 32 distal

and 32 proximal amputations. This is a valid comparison as it is known that proximal and distal blastemas go through the various stages of development at the same rate^{12,13}. The amount of hybridization varied from section to section because of variation in the number of blastema cells present in the sections irrespective of the level of amputation. When sections are viewed at a higher magnification so that hybridization to individual cells can be observed (Fig. 3c, d) there is no apparent difference in the amount of hybridization over individual cells between proximal and distal amputations. Thus, as with the anterior/posterior axis, there does not seem to be graded expression of the RAR along the proximal/distal axis.

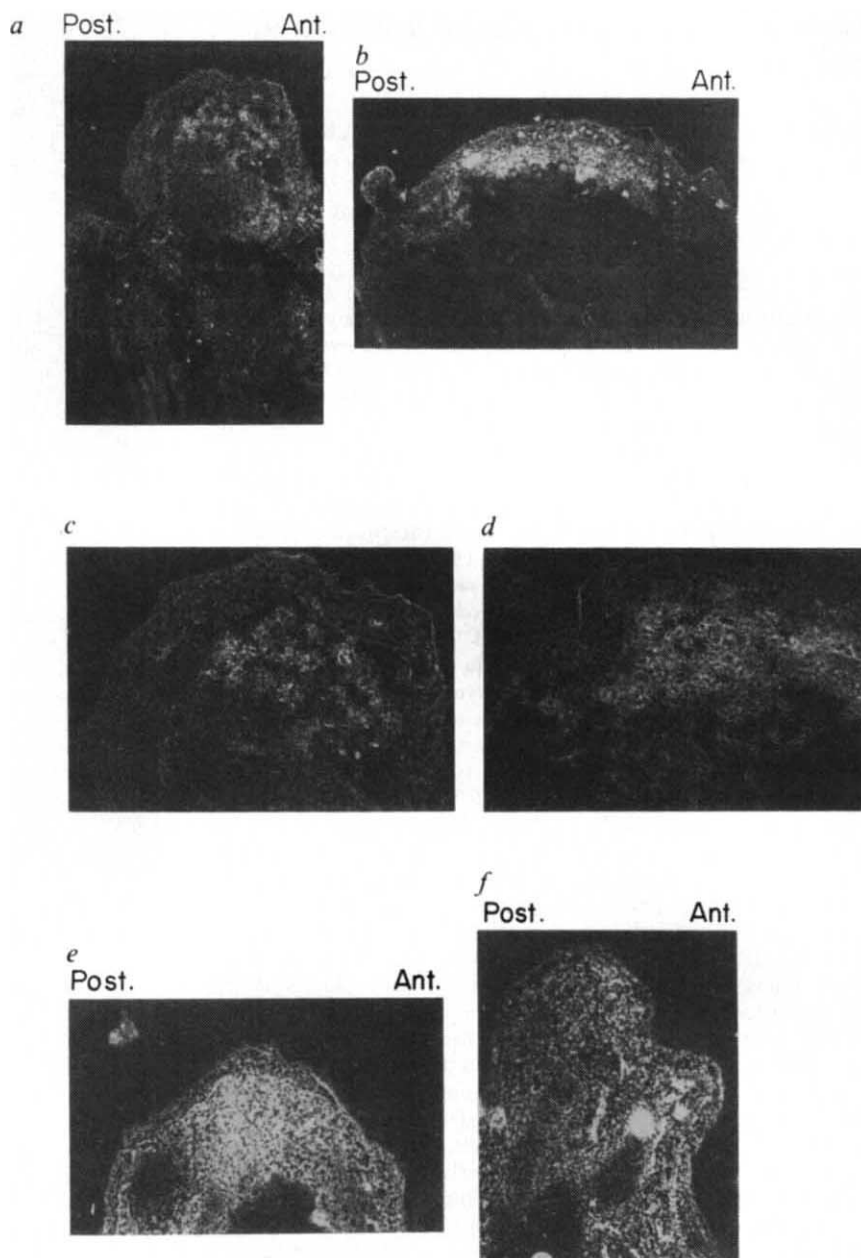
The optimal sensitivity of the 7-day regenerate to retinoic acid exposure suggests that loss of responsiveness may correspond to decreased or altered RAR expression. To examine this possibility, we looked at the expression of the RAR at later stages in regeneration. Sixteen days after amputation, cells of the medium bud blastema are morphologically uniform. Although significantly less responsive to retinoic acid, *in situ* hybridization reveals no decrease in RAR expression (Fig. 3e).

By day 25 after amputation the blastema has reached early digit stage. At this stage there has been considerable differentiation and exogenous retinoic acid can no longer alter the limb pattern (Table 1). In these late blastemas, *in situ* hybridization detects distinctly diminished RAR expression (Fig. 3f), either because of a decrease in the level of expression in each cell or a decrease in the percentage of cells expressing the RAR.

Although the effects of retinoids on limb development had been known for many years, it was the localization of endogenous retinoic acid in the limb buds that provided the crucial evidence that it may be a vertebrate morphogen¹⁴. This theory was advanced by the identification of the RAR whose

Fig. 3 Expression of the newt RAR β in regenerating limbs. *a, b*, Seven days post-amputation. *a* is from a proximal amputation (above the elbow) and *b* is from a distal amputation (through the wrist). Both were from left forelimbs, and anterior and posterior are marked (Ant. and Post.). *c, d*, Magnified comparison between expression of the RAR in proximal and distal blastemas. Higher magnification of 7-day blastema *in situ* hybridizations showing the signal to the RAR β probe on individual cells. *c* is from a proximal amputation and *d* is from a distal amputation (same exposure settings and developing times were used in printing *c* and *d*). *e, f*, Expression of the RAR in mid and late stage blastemas. *In situ* hybridizations to day 16 (mid-bud stage) and day 25 (early-digit stage) blastemas. *e* is from a mid-bud blastema and *f* is from an early digit blastema. Both were amputated through the mid humerus of forelimbs.

Methods. On the stated day after amputation, limbs were removed, fixed in 4% formaldehyde, washed in 30% sucrose overnight, embedded dorsal side up in OTC, frozen, and then 8 μ M cryostat sections were taken on subbed slides²⁰. After baking, the slides were hybridized to either antisense or sense (not shown) newt RAR ³⁵S-labelled RNA probes. The probes were made by cutting the plasmid pGEM NB8 with *Xho*I and transcribing either with SP6 (sense) or T7 (antisense) polymerase. The antisense probe was thus 102 nucleotide long, primarily corresponding to part of the retinoic acid-binding domain of the RAR. The probes were purified by PAGE. After washing, the slides were dipped in photographic emulsion and exposed for 10 days before developing in Kodak D9 developer. For all amputations, newts (*Notophthalmus viridescens* from Charles Sullivan, Inc.) were anaesthetized with 0.1% tricaine (Sigma). Limbs were cut with scissors, the bone(s) trimmed back as much as possible and excess skin removed. Under a microscope sulphamerizene (0.5%, Sigma) was used post-operatively for 24 hours to reduce infection.



homology to the steroid receptors has dramatic implications for how retinoids might act^{1,2}. This molecule is distinct from the retinoic acid-binding protein (CRABP) which displays graded expression in developing limbs and may be involved in amplifying the retinoic acid signal¹⁵. This would be consistent with our observations that RAR transcripts are homogeneously detected in regenerating cells shortly after amputation and continue to be expressed at a high level until differentiation and morphogenesis of the regenerated limb are complete. The presence of the RAR in the mesenchymal regenerating blastema cells at the time when retinoic acid has its maximal effect, provides strong evidence that retinoic acid establishes positional information through differential receptor activation of gene expression.

We thank Kazuhiko Umesono for plasmids and advice in the gene transfer studies, and members of the Evans lab for discussions throughout this work. We also wish to thank Elaine Stevens for preparation of the manuscript and expert secretarial assistance. V.G. is a Centennial fellow of the Medical Research Council of Canada. R.M.E. is an Investigator of the Howard Hughes Medical Institute. This work was supported by grants

from the NIH, the Howard Hughes Medical Institute, Mathers Foundation and Hoechst AG.

Received 25 October 1988; accepted 4 January 1989.

1. Giguère, V., Ong, E. S., Segui, P. & Evans, R. M. *Nature* **330**, 624-629 (1987).
2. Petkovich, M., Brand, N. J., Krust, A. & Chambon, P. *Nature* **330**, 444-450 (1987).
3. Evans, R. M. *Science* **240**, 889-895 (1988).
4. Robertson, M. *Nature* **330**, 420-421 (1987).
5. Brand, N. *et al.* *Nature* **332**, 850-853 (1988).
6. Benbrook, D., Lernhardt, E. & Pfahl, M. *Nature* **333**, 669-672 (1988).
7. Umesono, K., Giguère, V., Glass, C. K., Rosenfeld, M. G. & Evans, R. M. *Nature* **334**, 262-265 (1988).
8. Sporn, M. B. & Roberts, A. B. in *The Retinoids* (eds Sporn, M. B. *et al.*) 236-281 (Academic, Florida, 1984).
9. Niazi, I. A. & Saxena, S. A. *Folia Biol., Krakow* **26**, 3-8 (1978).
10. Maden, M. *J. Exp. Zool.* **230**, 387-392 (1984).
11. Kim, W.-S. & Stocum, D. L. *Dev. Biol.* **114**, 170-179 (1986).
12. Spallanzani, L. *Prodomo su un Opera da Imprimersi sopra le Riproduzioni Animali Modena* (1786).
13. Iten, L. F. & Bryant, S. V. *Wilhelm Roux Arch. dev. Biol.* **173**, 263-282 (1973).
14. Thaller, C. & Eichele, G. *Nature* **327**, 625-628 (1987).
15. Maden, M., Ong, D. E., Summerhill, D. & Chytil, F. *Nature* **335**, 733-735 (1988).
16. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
17. Giguère, V., Hollenberg, S. H., Rosenfeld, M. G. & Evans, R. M. *Cell* **46**, 645-652 (1986).
18. Glass, C. K., Holloway, J. M., Devari, O. & Rosenfeld, M. G. *Cell* **54**, 313-323 (1988).
19. de Wet, J. R., Wood, K. V., deLuca, M., Helinski, D. R. & Subramani, S. *Molec. cell Biol.* **7**, 725-737 (1987).
20. Hogan, B., Constantini, F. & Lacy, E. *Manipulating the Mouse Embryo: A Laboratory Manual* 228-242 (Cold Spring Harbor, New York, 1986).

Localization of muscle gene products in nuclear domains

Grace K. Pavlath*, Kevin Rich*, Steven G. Webster & Helen M. Blau†

Department of Pharmacology, Stanford University School of Medicine, Stanford, California 94305-5332, USA

The localization of gene products is central to the development of cell polarity^{1,2} and pattern specification during embryogenesis³⁻⁵. To monitor the distribution of gene products encoded by different nuclei in the same cell in tissue culture, we fused cells of different species to form multinucleated non-dividing heterokaryons^{6,7}. In previous fusion studies, cell-surface antigens and organelles contributed by disparate cell types intermixed within minutes⁸⁻¹¹. Using heterokaryons produced with differentiated muscle cells, we demonstrate here that a muscle membrane component, the Golgi apparatus mediating its transport, and a sarcomeric myosin heavy chain are localized in the vicinity of the nuclei responsible for their synthesis. These results provide direct evidence that products (organelle, membrane and structural proteins) derived from individual nuclei can remain localized in myotubes, a finding with implications both for neuromuscular synapse formation and for the carrier state of Duchenne muscular dystrophy.

To examine the distribution of gene products derived from different nuclei, we produced two types of heterokaryons. Differentiated mouse muscle cells (C2C12) were fused with human non-muscle cells (fibroblasts, keratinocytes and hepatocytes) using polyethylene glycol (PEG) as previously described^{6,7}. In addition, mouse myoblasts and human myoblasts were plated together and allowed to fuse spontaneously. Fusion products derived from two different cell types were identified by the punctate (mouse) and uniform (human) nuclear-staining pattern obtained with Hoechst 33258. After fusion the location of human muscle products was distinguished in single heterokaryons using species-specific antibodies and immunofluorescence microscopy.

5.1H11 is a human muscle membrane protein expressed at low levels on mononucleated myoblasts and at high levels on multinucleated differentiated myotubes, but not detected on non-muscle cells cultured in the same dish. Expression of 5.1H11 is induced in a number of human non-muscle cell types after fusion with mouse muscle cells⁷. Two patterns of distribution of 5.1H11 were apparent on the membranes of PEG-fused and spontaneous heterokaryons: in some cases the antigen was distributed along the entire length of the cell (Fig. 1a), but it was usually localized in the vicinity of the human nucleus responsible for its synthesis (Fig. 1b-d). The dimensions of such nuclear domains varied, possibly as a result of nuclear movement¹². This unequal distribution of a cell-surface antigen contrasts with earlier reports that antigens rapidly intermix and remain dispersed⁸⁻¹⁰ after cell fusion.

A sarcomeric myosin heavy chain (SMHC), a structural protein with a central role in muscle contraction, was also localized in the vicinity of the nucleus responsible for its production in heterokaryons. To monitor human SMHC expression, we used a monoclonal antibody against slow SMHC (4A.951) (ref. 13), an isoform that is synthesized by human but not by mouse C2C12 muscle cells. In heterokaryons, slow SMHC was detected (Fig. 2). This SMHC isoform presumably derived from human fibroblast nuclei which carry out a human myogenic programme in response to mouse muscle regulators in heterokaryons¹⁴. Moreover, although SMHCs were distributed throughout

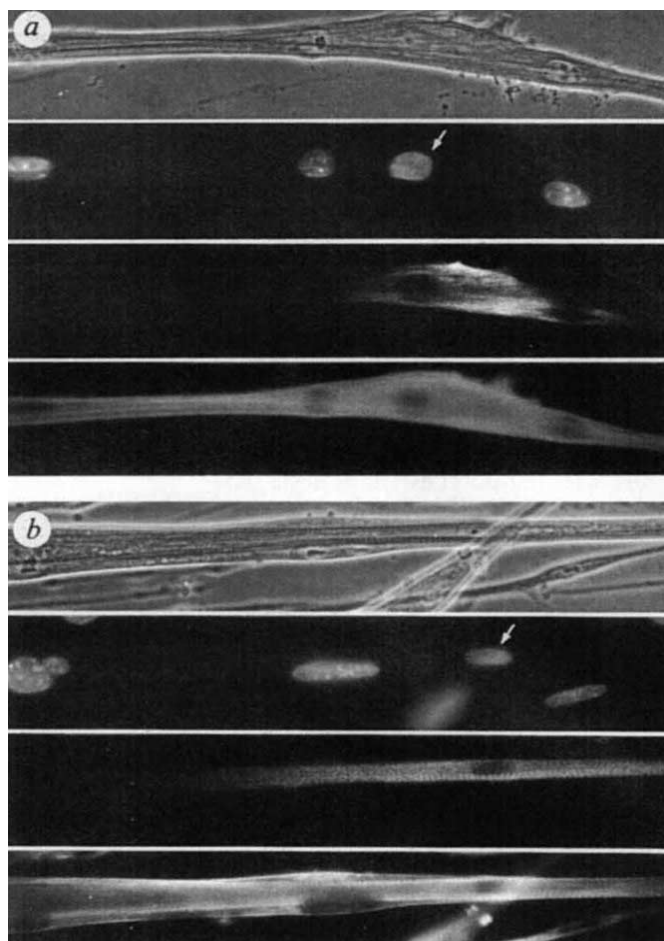


Fig. 2 Localized distribution of myosin heavy chain in heterokaryons. Two heterokaryons (*a* and *b*) are shown with myosin heavy chain localized in the vicinity of the human nucleus (arrows) responsible for encoding it. In each case, the same field is shown in four views, from top to bottom: phase contrast to define myotube boundaries, Hoechst 33258 to distinguish mouse and human nuclei, monoclonal antibody 4A.951 to reveal human slow SMHC isoforms, and monoclonal antibody 4A.1025 to reveal all SMHC isoforms. The heterokaryon in (*a*) was formed by PEG fusion and analysed 4 days after fusion, whereas the heterokaryon in (*b*) was formed by spontaneous fusion and analysed 7 days after co-culture. Cell types are described in Fig. 1. Partial views of heterokaryons are shown; total numbers of mouse and human nuclei (mouse, human) in these heterokaryons are: *a*, 5,2 and *b*, 7,1. Cells were fixed as previously described¹³, then reacted in four sequential steps with 4A.951 hybridoma supernatant for 90 min, with fluorescein isothiocyanate-conjugated goat anti-mouse IgG (1:100) for 60 min, then with biotinylated monoclonal antibody 4A.1025 (1:200) for 45 min, Texas red-avidin (1:800) for 15 min, then stained with Hoechst 33258 and visualized as previously described^{7,13}.

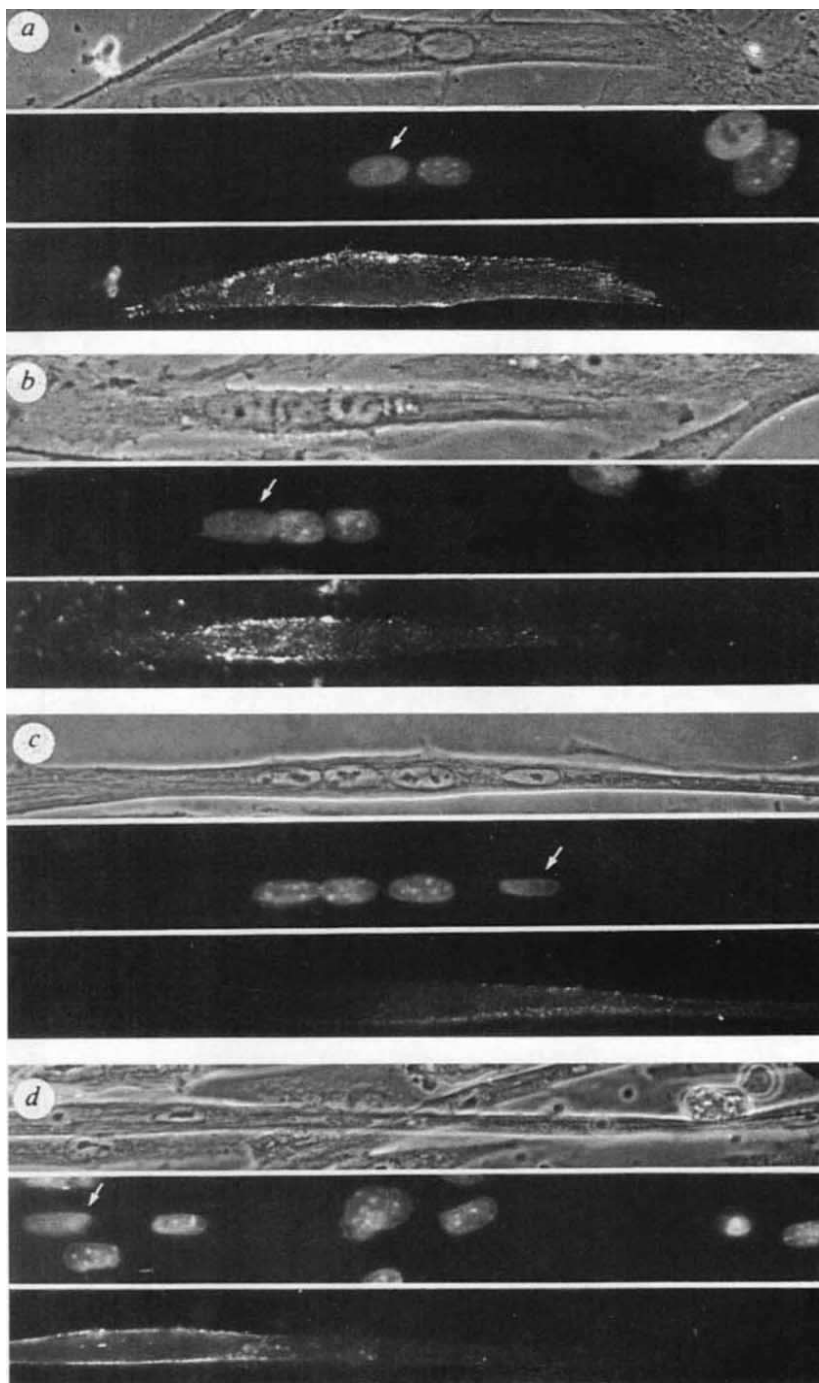
heterokaryons, the human slow SMHC isoform was most frequently found near human nuclei.

The probability of a heterokaryon displaying a localized distribution of the cell-surface or sarcomeric gene products was directly correlated with the proportion of muscle nuclei it contained. At a ratio of muscle to fibroblast nuclei of $\geq 4:1$, localization was observed in $89 \pm 3\%$ and $84 \pm 6\%$ of all heterokaryons that expressed 5.1H11 and SMHC respectively (Fig. 3, left). This finding was not specific to fibroblasts, which have extensive stress fibres that might impair intracellular movement of proteins¹⁵. Similar results were obtained with heterokaryons produced with other human non-muscle cell types (hepatocytes and keratinocytes), although the absolute frequency with which localization occurred differed.

* Present addresses: Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona 85721, USA (G.K.P.) and Smithsonian Institution, Washington, DC, USA (K.R.).

† To whom correspondence should be addressed.

Fig. 1 Localized distribution of 5.1H11 antigen on the surface of heterokaryons. *a*, Heterokaryon with a non-localized distribution of 5.1H11. *b-d*, Three heterokaryons with the cell-surface antigen 5.1H11 localized in the vicinity of the human nucleus (arrows) encoding it. Triplicate panels for the four different heterokaryons (*a-d*) show boundaries of individual multinucleated cells in phase contrast (top), nuclei stained with Hoechst 33258 (middle) to distinguish uniformly stained human nuclei (arrows) from punctate mouse nuclei, and distribution of 5.1H11 human muscle antigen by immunofluorescence (bottom). The 5.1H11 antigen was visualized by reaction with monoclonal antibodies in a hybridoma supernatant at 37°C followed by biotin-conjugated anti-mouse IgG and finally avidin-conjugated Texas Red, as previously described⁷. Owing to their large size, only portions of heterokaryons are shown. Total numbers of mouse and human nuclei contained in these heterokaryons are (mouse, human): *a*, 1,1; *b*, 2,1; *c*, 4,3 and *d*, 8,1, analysed 4 days (*a-c*) and 7 days (*d*) after fusion. Heterokaryons (*a-c*) were produced by PEG-fusion of mouse C2C12 myotubes and human MRC-5 fetal lung fibroblasts, as previously described^{6,7}. The heterokaryon in (*d*) was formed by the spontaneous fusion of mouse C2C12 myoblasts cultured with human primary myoblasts, which had been enriched to 99% homogeneity with the fluorescence-activated cell sorter using monoclonal antibody to 5.1H11 (ref. 27).



The protein products derived from individual nuclei might be expected to localize initially, but later to encompass the entire cell as the proteins diffuse⁸⁻¹⁰. But a time course revealed that the proportion of heterokaryons with a localized distribution of 5.1H11 and SMHC was constant over a 7-day period of time for each ratio of muscle to non-muscle nuclei (Fig. 3, right). Thus, a localized distribution does not appear to precede a uniform distribution, but is correlated with muscle nuclear input.

The finding that in spontaneous heterokaryons 5.1H11 and SMHC are also localized near the nucleus containing the gene that encoded them (Fig. 1*d*; Fig. 2*b*; and Fig. 3, bar S) indicates that nuclear domains could be an intrinsic property of muscle cells. Thus, localization is not due to perturbation of cell membranes by PEG, to the aberrant combination of two disparate cell types in one cell, or to the fusion of a non-muscle cell into a pre-existing well-developed muscle structure.

The Golgi apparatus which mediates the intracellular trans-

port of the cell-surface antigen 5.1H11 is also localized within nuclear domains (Fig. 4). To visualize the human Golgi, an antibody to the human Golgi enzyme galactosyl transferase was used at a concentration at which it was species-specific (Fig. 4*a*). Typically, when the nuclei were closely juxtaposed, the Golgi were shared (Fig. 4*b* and *c*, left), whereas when the nuclei were separate, the Golgi were distinct (Fig. 4*c*, right and Fig. 4*d*), indicating that the organelles and secretory pathways of neighbouring nuclei in muscle cells, unlike other hybrid cells¹¹, can remain discrete.

The 5.1H11 cell-surface antigen, the Golgi apparatus and SMHC examined here are retained in regions near their site of synthesis in large multinucleated myotubes. But the movement of some muscle gene products is not limited. In heterokaryons mouse *trans*-acting regulatory factors gain access to and activate human nuclei, and subunits of the enzyme creatine kinase form mouse-human heterodimers more frequently than human

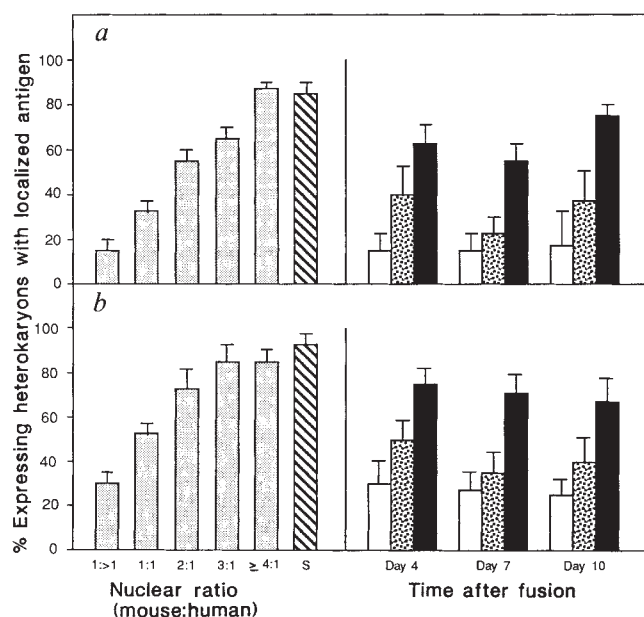


Fig. 3 Localized 5.1H11 (a) and myosin heavy chain (b), in heterokaryons is correlated with an increased proportion of muscle nuclei (left) and is constant over time (right). The frequency of localization was scored as a function of nuclear ratio (a) over time (b) for nuclear ratios of mouse to human of 1:>1 (white), 1:1 (stippled), and >1:1 (black). Heterokaryons that formed spontaneously (S) when mouse and human myoblasts were cocultured also showed localization (striped). In a and b a total of 936 heterokaryons that had an average of 7 nuclei were analysed (range 2 to 97 nuclei). Of these, $87 \pm 2\%$ expressed 5.1H11 and $49 \pm 2\%$ expressed SMHC. Nuclear ratio, 5.1H11 and SMHC 4A.951 were assayed as described in Figs 1 and 2. The results are representative of four different time course experiments. Errors represent the standard error of the proportion⁷. No significant differences in the frequency of localization were observed when heterokaryons were analysed in a separate time course at 12-hour intervals spanning 1 to 4 days after fusion (data not shown). The frequency of localization in heterokaryons (nuclear ratio of 3:1) containing five different cell types differed: $77 \pm 5\%$ for fetal lung fibroblasts, $69 \pm 13\%$ for hepatocytes, $31 \pm 8\%$ for keratinocytes and $29 \pm 5\%$ for fetal and adult skin fibroblasts.

homodimers⁶. In addition, a proportion of acetylcholine receptors is mobile¹⁶, and unlike 5.1H11, is rapidly internalized when live myotubes are exposed to antibodies (ref. 17; Kaplan and H.M.B., unpublished observations).

A variety of mechanisms could mediate the localization of muscle proteins. In the case of SMHC, localization may reflect interactions necessary for the targeted insertion of SMHC isoforms during sarcomere assembly^{18,19}. The localization of cell-surface antigens is likely to reflect not only targeting, but also mechanisms that restrict diffusion within myotube membranes. We find that the mobility of the 5.1H11 antigen is markedly reduced during muscle differentiation: antibodies induce 5.1H11 to cluster into small patches on myoblasts but have minimal effects on myotubes (K.R. and H.M.B., unpublished observations). Thus, the localization of cell-surface antigens could be due to altered membrane composition²⁰ or to specific cytoskeletal interactions²¹.

It has been suggested that the acetylcholine receptors concentrated at neuromuscular synapses may be derived specifically from muscle nuclei in the immediate vicinity of the neuromuscular junction^{22,23}. Our findings provide direct evidence that such a localization of products derived from individual nuclei can occur. Nuclear domains could also be the basis for the distinct SMHC isoforms produced in muscle fibres at sites of ectopic innervation²⁴. In addition, our results suggest why as many as two-thirds of carriers for Duchenne muscular dystrophy

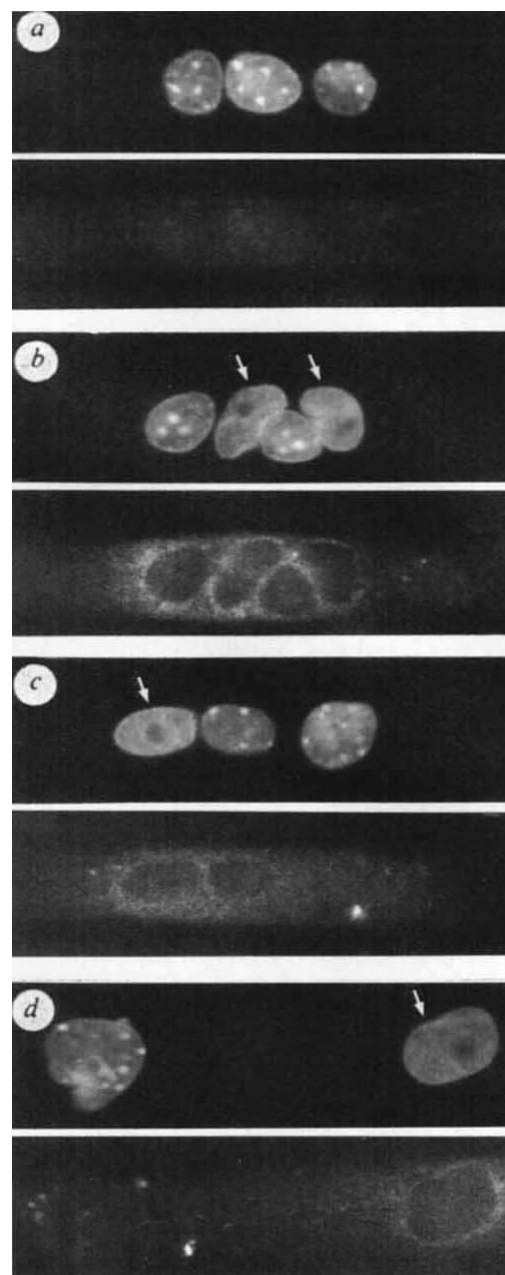


Fig. 4 Localized distribution of Golgi apparatus in heterokaryons. a, An affinity-purified polyclonal antibody (O₁₂) to the human Golgi enzyme galactosyl transferase²⁸ is a human-specific reagent at a concentration of 1:10 and does not react with the mouse Golgi enzyme in a myotube containing only mouse muscle nuclei. In b and c (left), when human (arrows) and mouse nuclei are closely juxtaposed, the human Golgi is shared. In c, (right) and d, when human and mouse nuclei are separated, the human Golgi is not shared and remains distinct. In contrast to the Golgi of most cell types, the Golgi shown here is circumnuclear, a distribution typical of striated muscle cells^{29,30} that is induced in non-muscle cells after fusion with muscle cells to form heterokaryons³¹. Protocols for fixation and immunofluorescence staining were as previously described³¹.

exhibit muscle defects²⁵, although their muscle cells are natural heterokaryons for X-chromosome-linked functions and complementation might be expected to mask the effects of the defective gene product. The gene product, which appears to be a cytoskeletal or membrane component²⁶, may be localized in nuclear domains, resulting in focal lesions along a fibre leading to elevated serum creatine kinase. Why muscle cells are com-

posed of nuclear domains remains unknown. Possibly the localization of muscle gene products in multinucleated cells, like the localization of transcripts in early embryos³⁻⁵, provides a means for maintaining positional information in large cells.

We thank Drs E. Berger and F. Walsh for Golgi and 5.1H11 antibodies, respectively, Drs Z. Hall, L. Rubin, J. Singer and B. Wallace for criticism of the manuscript, M. Travis for technical assistance and K. Bird for secretarial help. This work was supported by a predoctoral training grant (G.K.P. and S.G.W.) and funds from the National Institutes of Health, the National Science Foundation and the Muscular Dystrophy Association of America. H.M.B. is the recipient of a Research Career Development Award.

Received 31 October; accepted 16 December 1988.

- Misek, D. E., Bard, E. & Rodriguez-Boulant, E. *Cell* **39**, 537-546 (1984).
- Segev, N., Mulholland, J. & Botstein D. *Cell* **52**, 915-924 (1988).
- Gurdon, J. B., Mohun, T. J., Fairman, S. & Brennan, S. *Proc. natn. Acad. Sci. U.S.A.* **82**, 139-143 (1985).
- Melton, D. A. *Nature* **328**, 80-82 (1987).
- Driever, W. & Nüsslein-Volhard, C. *Cell* **54**, 83-93 (1988).
- Blau, H. M., Chiu, C.-P. & Webster, C. *Cell* **32**, 1171-1180 (1983).
- Blau, H. M. *et al. Science* **230**, 758-766 (1985).
- Watkins, J. F. & Grace, D. M. *J. Cell Sci.* **2**, 193-204 (1967).
- Harris, H., Sidebottom, E., Grace, D. M. & Bramwell, M. E. *J. Cell Sci.* **4**, 499-525 (1969).
- Frye, L. D. & Edidin, M. *J. Cell Sci.* **7**, 319-335 (1970).
- Tassin, A.-M., Maro, B. & Bornens, M. *J. Cell Biol.* **100**, 35-46 (1985).
- Englander, L. L. & Rubin, L. L. *J. Cell Biol.* **104**, 87-95 (1987).
- Webster, C., Silberstein, L., Hays, A. P. & Blau, H. M. *Cell* **52**, 503-513 (1988).
- Hardeman, E., Chiu, C.-P., Minty, A. & Blau, H. M. *Cell* **47**, 123-130 (1986).
- Ash, J. F., Louvard, D. & Singer, S. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5584-5588 (1977).
- Stya, M. & Axelrod, D. *Proc. natn. Acad. Sci. U.S.A.* **80**, 449-453 (1983).
- Bursztajn, S., McManaman, J. L., Elias, S. B. & Appel, S. H. *Science* **219**, 195-196 (1983).
- Isaacs, W. B. & Fulton, A. B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6174-6178 (1987).
- Taylor, L. D. & Bandman, E. *J. Cell Biol.* (in the press).
- Kawata, S., Lehner, C., Müller, M. & Cherry, R. C. *J. biol. Chem.* **257**, 6470-6476 (1982).
- Edelman, G. M. *Science* **192**, 218-226 (1976).
- Merrill, J. P. & Sanes, J. R. *Nature* **317**, 66-68 (1985).
- Fontaine, B., Sassoon, D., Buckingham, M. & Changeux, J.-P. *EMBO J.* **7**, 603-609 (1988).
- Salviati, G., Biasia, E. & Aloisi, M. *Nature* **322**, 637-639 (1986).
- Hausmanowa-Petrusewicz, I., Niebroj-Dobosz, I., Borkowska, J., Lukasik, E. & Liszewska-Pfeiffer, D. in *Pathogenesis of Human Muscular Dystrophies* (ed. Rowland, L. P.) (Excerpta Medica, Oxford, 1977).
- Hoffman, E. P., Brown, R. H. & Kunkel, L. M. *Cell* **51**, 919-928 (1987).
- Webster, C., Pavlath, G. K., Parks, D. R., Walsh, F. S. & Blau, H. M. *Exp. Cell Res.* **174**, 252-265 (1988).
- Berger, E. G., Mandel, T. & Schilt, U. *J. Histochem. Cytochem.* **29**, 364-370 (1981).
- Tassin, A.-M., Paintrand, M., Berger, E. G. & Bornens, M. *J. Cell Biol.* **101**, 630-638 (1985).
- Kronebusch, P. J. & Singer, S. J. *Cell* **88**, 25-34 (1987).
- Miller, S. C., Pavlath, G. K., Blakely, B. T. & Blau, H. M. *Genes Dev.* **2**, 330-340 (1988).

Functional cooperativity between protein molecules bound at two distinct sequence elements of the immunoglobulin heavy-chain promoter

Lorenz Poellinger, Barbara K. Yoza & Robert G. Roeder

Laboratory of Biochemistry and Molecular Biology,
The Rockefeller University, 1230 York Avenue, New York,
New York 10021, USA

Immunoglobulin heavy-chain gene promoters contain two conserved upstream sequence elements, octamer and heptamer¹⁻⁴, both of which are required for normal cell type-specific promoter function *in vivo*⁴⁻⁶. The octamer sequence motif 5'-ATGCAAT-3', and its precise inverse, are strongly conserved in heavy- and light-chain gene promoters^{1,2} and are important determinants for the lymphoid-specific function of these promoters⁷⁻⁹ and of the heavy-chain enhancer^{10,11}. The heptameric sequence element with the consensus 5'-CTCATGA-3' (refs 3 and 4) is also required in addition to the octamer for full lymphoid-specific activity of heavy-chain promoters^{4,6}. Although these two elements have no sequence similarity, they are both recognized *in vitro*¹² by the ubiquitous octamer transcription factor OTF-1 (reviewed in refs 13 and 14) and the lymphoid-specific OTF-2 (reviewed in refs 15 and 16).

Here we show that purified OTF-2 binds cooperatively to the immunoglobulin heptamer and octamer elements so that interaction with the octamer element facilitates binding of OTF-2 to the heptamer motif. More important, cooperativity in OTF-2 binding is closely mirrored by functional cooperation between the heptamer and octamer elements in activating transcription from the heavy-chain promoter *in vitro*.

The lymphoid-specific transcriptional activator protein OTF-2, the first mammalian cell type-specific gene regulatory factor identified^{17,18}, has been purified to apparent homogeneity in a transcriptionally active form¹⁵ and structurally characterized by complementary DNA cloning^{16,19}. We have previously shown that OTF-1 and OTF-2 can bind to the heptamer and octamer motifs of the BCL1 immunoglobulin heavy-chain promoter to form dimer complexes¹². Here we establish a cooperativity in the binding of purified OTF-2 to the BCL1 heptamer and octamer and demonstrate that these two elements cooperate to activate transcription from the heavy-chain promoter *in vitro*.

A gel mobility shift assay was used to characterize the interaction of the purified¹⁵ OTF-2 protein (relative molecular mass, 60,000) from Namalwa cells (a human Burkitt lymphoma line) with the heavy-chain gene heptamer and octamer sequence elements. Purified OTF-2 generated two distinct protein-DNA complexes (C1 and C2 in lane 2 of Fig. 1a) in binding experiments with a wild-type oligonucleotide probe (H⁺O⁺) containing the heptamer and octamer elements of the BCL1 heavy-chain promoter¹⁷, in agreement with previous observations.¹² Formation of complex C2 must be dependent on the presence of an intact heptamer motif, because it was not observed with an oligonucleotide probe (H⁺O⁺) containing clustered point mutations in the heptamer¹² (Fig. 1a; compare lanes 2 and 4). Analysis at a fixed concentration of the H⁺O⁺ template showed that low concentrations of OTF-2 generated predominantly one complex (C1), corresponding to occupation of the octamer element, whereas higher concentrations resulted in formation of the heptamer-dependent complex (C2) as well (Fig. 1b, upper panel). Quantitation by scintillation counting of the OTF-2-induced protein-DNA complexes confirmed that complex C1 was generated before formation of complex C2 (Fig. 1b, lower panel). Formation of complex C2 followed sigmoidal equilibrium binding kinetics, indicating that OTF-2 interaction with the heptamer is facilitated by prior occupation of the octamer element by OTF-2.

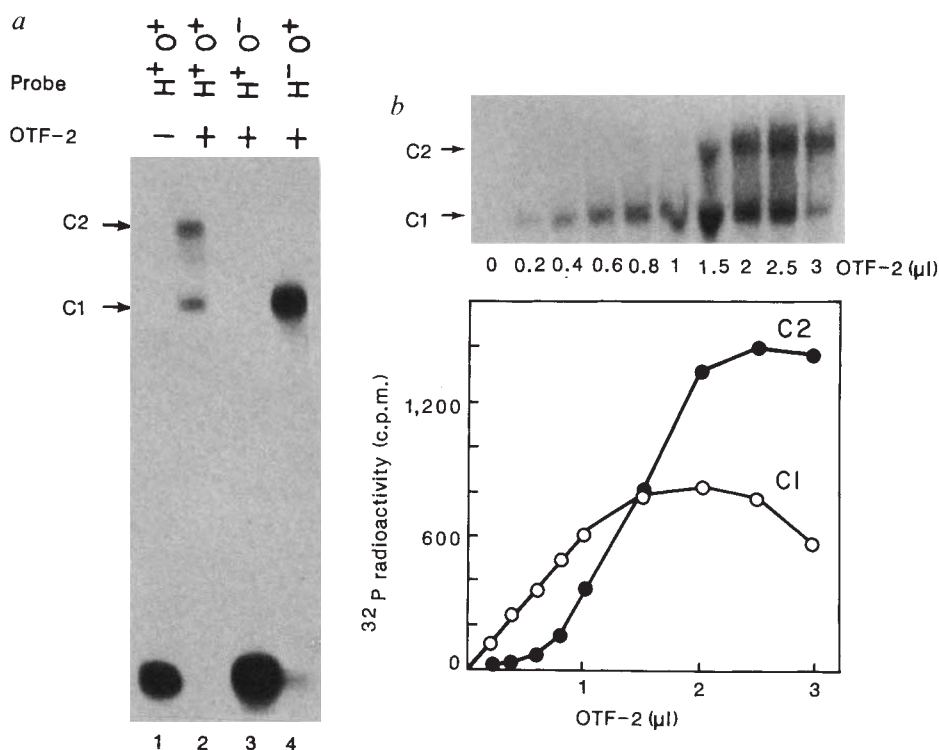
Using the gel mobility shift assay, we next compared the ability of unlabelled wild-type H⁺O⁺ and mutant H⁺O⁺ or H⁺O⁻ oligonucleotides to compete for the binding of OTF-2 to the octamer motif of the human histone H2B promoter¹³. The wild-type H⁺O⁺ DNA fragment competed very effectively for OTF-2 against the histone H2B octamer (Fig. 2a), whereas comparable levels of competition (for example, reduction to 50%) by the mutant H⁺O⁺ DNA fragment required a roughly fourfold increase in molar amount (Fig. 2b). At these competitor concentrations, neither a DNA fragment (μ E3, ref. 20) that contained unrelated sequences from the heavy-chain enhancer, nor the mutant H⁺O⁻ DNA fragment competed for binding (Fig. 2). Thus, the heptamer and octamer elements seem to act together to affect the affinity of the heavy-chain promoter fragments for OTF-2.

To investigate the functional relevance of this heptamer-dependent cooperative binding of OTF-2, wild-type or mutant oligonucleotides containing the heptamer and octamer elements and the TATA box of the BCL1 heavy-chain promoter (Fig. 3a) were inserted in front of a rabbit β -globin test-gene construct²¹. Thus the globin ATA box was replaced by the heavy-chain TATA box, but the distance between the TATA box and transcription initiation site was identical to that in the BCL1 promoter¹⁷. The plasmid containing the wild-type H⁺O⁺ heavy-chain motif gave high levels of accurate transcription initiation from circular templates over a range of concentrations of Namalwa nuclear extract (Fig. 3b). This level of transcription

Fig. 1 Cooperative binding of purified OTF-2 to the immunoglobulin heptamer and octamer sequence elements. *a*, Autoradiograph showing gel mobility shift assay of purified OTF-2 using the wild-type (H^+O^+), octamer-mutant (H^+O^-) or heptamer-mutant (H^-O^+) oligonucleotide probes. Protein-DNA complexes C1 and C2 are indicated by arrows; unbound probe runs at the bottom of the autoradiograph. *b*, Saturation binding studies. Increasing amounts of purified OTF-2 were incubated with the wild-type H^+O^+ oligonucleotide probe and assayed by the gel mobility shift assay. The upper panel shows the autoradiograph with protein-DNA complexes C1 and C2 indicated by arrows. The lower panel shows the equilibrium binding curve for complexes 1 (open circles) and 2 (closed circles) after quantitation by liquid scintillation spectrometry.

Methods. OTF-2 was purified from Namalwa cell nuclear extracts³¹ essentially as described¹⁵, except that a single-stranded DNA-agarose column was used for non-specific DNA-affinity chromatography.

Purified OTF-2 was used for the gel mobility shift analysis using ~4 fmol of radiolabelled probe. The gel electrophoresis DNA-binding assay was essentially as described¹⁵. Shifted bands were excised from the gel and assayed for radioactivity for quantitation. The wild-type H^+O^+ oligonucleotide probe spans the heptamer and octamer elements of the BCL1 heavy-chain promoter¹⁷ extending from position -73 to position -42 from the initiation site. The clustered point mutations in either the octamer or heptamer elements are identical to those shown in Fig. 3a.



was consistently decreased to about half when the heptamer sequence was altered (Fig. 3b; compare lanes 1, 4 and 7 with lanes 2, 5 and 8, respectively). In contrast, transcription from the octamer-defective H^+O^- promoter resulted in very low (presumably basal) activity (<10% of wild-type signal; Fig. 3b, lanes 3, 6 and 9). Moreover, the relative differences in transcriptional activity between the wild-type promoter and the H^+O^+ and H^+O^- mutants were maintained in HeLa nuclear extracts, although the overall levels of transcription were markedly lower. Thus, wild-type promoter activity in HeLa nuclear extracts corresponded roughly to the very low octamer-minus promoter activity observed in Namalwa extract (Fig. 3b; compare lanes 1, 4 and 7 with lanes 10, 13 and 16 respectively). The low transcription from the wild-type promoter in HeLa cells presumably reflects some activity of OTF-1 towards this template *in vitro*. As a control, the adenovirus-type 2 major late promoter was transcribed using both extracts with very similar efficiency (Fig. 3c). It is important that in Namalwa extracts transcription from the wild-type H^+O^+ promoter (Fig. 4) or the mutant H^+O^- promoter (data not shown) showed the same relative sensitivity to competitor oligonucleotides as did the specific DNA-binding activity of purified OTF-2. When added to nuclear extracts programmed with the wild-type promoter, the H^+O^+ oligonucleotide inhibited transcription ~4 times more efficiently than the H^+O^- template oligonucleotide. Finally, the octamer-mutant H^+O^- oligonucleotide was inactive as a competitor in the concentration range tested (Fig. 4), but an approximately fivefold increase in competitor H^+O^+ DNA was required to abolish transcription, as compared with that needed to abolish protein binding. Presumably this difference reflects the presence of nonspecific DNA-binding proteins in the transcription assays.

In summary, *in vitro* transcription of point mutants in the heptamer and octamer elements shows that both are required for maximal transcription and that they have a synergistic effect on promoter activity. Whereas the octamer element alone can elicit about half the maximal transcriptional effect *in vitro*, the heptamer alone is inactive. The transcription template contained

no known promoter elements other than the heavy-chain heptamer and octamer elements and the TATA box. Thus, in agreement with recent *in vivo* observations^{8,9}, our *in vitro* results indicate that the octamer transcription factor(s) does not need to interact with any other upstream binding factors to stimulate transcription efficiently from a target promoter in a cell type-specific manner. As the difference in OTF content in Namalwa and HeLa cells is at most 2-3-fold^{10,15}, the observed tissue-specificity for the high-level transcription of the heavy-chain promoter does not seem to be attributable to quantitative differences in total amounts of OTF protein, as has recently been suggested²².

It is not clear how the cooperative effect between the heptamer and octamer elements and their bound OTF molecules is achieved. Our binding data support a model in which cooperativity may be established by direct contact between the bound OTF proteins, as illustrated already by the cooperativity of the phage lambda repressor bound to operator sites²³. Functional cooperativity has been demonstrated in several eukaryotic systems (reviewed in refs 24 and 25) but has rarely been substantiated at the DNA-binding level as it has in prokaryotes. It is probable that protein-protein interaction is the basis for the synergism in binding and function between the TATA box-binding protein TFIID and upstream factors²⁶⁻²⁸. In the case of the OTFs, there is strong evidence for physical interaction between the proteins in that they seem to form heptamer-dependent dimer complexes when bound to the BCL1 H^+O^+ target DNA sequence, and also it is possible to generate heteromeric complexes using mixtures of expressed truncated forms of OTF-2 and the genuine 60 K cellular protein¹². Dimerization of OTF-2 takes place both at the naturally occurring 2-base pair and 14-base pair spacings between the heptamer and octamer elements¹² and may involve the recently identified 'leucine zipper'²⁹ structural motif of OTF-2 (refs 16 and 30). We do not yet know whether the observed cooperativity is dependent on strict helix alignment of the two OTF-binding elements. Finally, the cooperative occupation of the heptamer and octamer elements could render heavy-chain

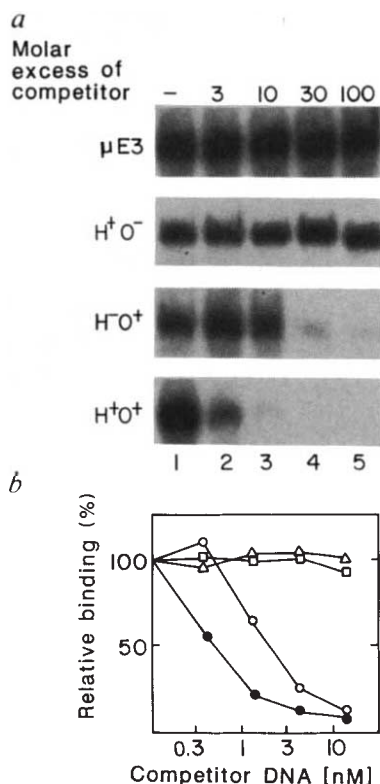


Fig. 2 Relative affinity of purified OTF-2 for wild-type or heptamer- or octamer-mutated oligonucleotide probes. The relative affinity of wild-type H⁺O⁺ or mutant H⁻O⁺ or H⁺O⁻ heavy-chain motifs for purified OTF-2 was determined by the gel mobility shift DNA-binding assay. The labelled template was a histone H2B octamer oligonucleotide probe¹². **a**, Gel mobility shift analysis showing competition of heavy-chain promoter oligonucleotides for H2B octamer-OTF-2 complexes. Lane 1, no competitor; lanes 2-5, competitor DNA added in increasing molar excess; panels from top to bottom show heavy-chain enhancer DNA (μE3; ref. 20), H⁺O⁻ mutant oligonucleotide, H⁻O⁺ mutant oligonucleotide and H⁺O⁺ wild-type oligonucleotide. **b**, Comparison of the relative affinity of the various heavy-chain motifs for OTF-2 after quantitation using the bound fragment. The concentrations of unlabelled H⁺O⁺ (closed circles), H⁻O⁺ (open circles) and H⁺O⁻ (triangles) or non-related μE3 (open squares) competitor DNA are indicated on the abscissa. Residual binding is indicated on the ordinate. The amount of radioactive DNA bound to OTF-2 in the absence of competitor was taken as 100%.

Methods. A fixed concentration (1 ng) of purified OTF-2 was incubated with 2 fmol of ³²P-labelled histone H2B octamer probe in the absence or presence of the indicated concentration of specific or nonspecific competitor DNA for 30 min at 30 °C. Protein-bound DNA was separated from free DNA on native polyacrylamide gels as shown in Fig. 1. The extent of binding was determined by excision of the bands representing bound and free DNA followed by scintillation counting.

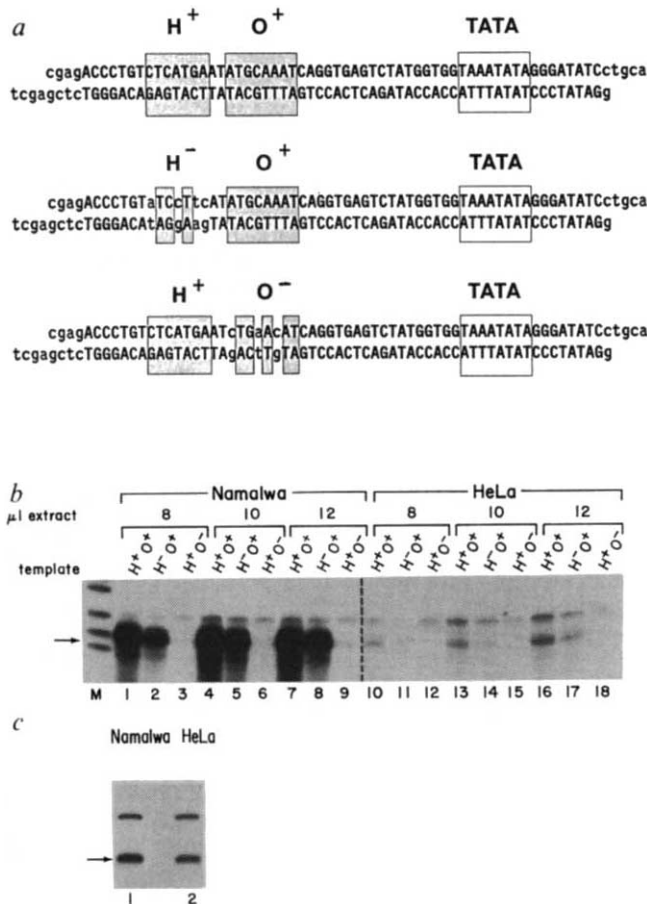


Fig. 3 **a**, Sequences of the wild-type (H⁺O⁺), heptamer-mutant (H⁻O⁺) or octamer-mutant (H⁺O⁻) oligonucleotides inserted into the *SacI*/*PstI* sites of the OVEC transcription vector²¹, immediately upstream of the β-globin transcription initiation site. The sequence of the H⁺O⁺ oligonucleotide corresponds to positions -73 to -16 of the BCL1 immunoglobulin heavy-chain promoter¹⁷. Restriction enzyme linker sequences and mutated nucleotides in the heptamer and octamer elements are indicated by lower-case letters. **b**, Transcription of the wild-type (H⁺O⁺), heptamer-mutant (H⁻O⁺), and octamer-mutant (H⁺O⁻) promoters in reaction mixtures containing the indicated amounts of Namalwa (lanes 1-9) and HeLa (lanes 10-18) cell nuclear extracts. The arrow indicates the accurately initiated transcription product. **c**, Comparison of transcription of the adenovirus major late promoter in Namalwa or HeLa cell nuclear extracts. The position of the accurately initiated transcript is indicated.

Methods. Namalwa and HeLa cells were grown as previously described¹⁵ and nuclear extracts prepared³¹. A standard transcription reaction contained 50 mM Tris pH 8.0, 12% glycerol, 46 mM KCl, 0.12 mM EDTA, 0.3 mM dithiothreitol, 0.16 mM PMSF, 3 mM MgCl₂, 600 μM each of ribonucleotide triphosphate, 10 μg template per ml and indicated amounts of nuclear extract at a concentration of 7.5 mg protein per ml. After incubation at 30 °C for 40 min, the reactions were stopped by the addition of an equal volume of 20 mM Tris pH 8.0, 0.6 M sodium acetate, 1% SDS, 20 mM EDTA and 200 μg yeast tRNA per ml. Reactions were extracted twice with phenol/CHCl₃, ether-extracted and ethanol-precipitated. RNA products were hybridized to a single-stranded, kinase-labelled probe for at least 18 h at 44 °C in 80% formamide and subsequently analysed by S1-nuclease protection³². The probe used in this analysis was a 90-base oligonucleotide containing sequences from -18 to +75 of the rabbit β-globin gene in the OVEC plasmid²¹. The adenovirus-type 2 major late promoter plasmid contained sequences from -51 to +192 cloned into the *AccI* and *HindIII* sites of pUC12. Transcription reactions using 10 μl nuclear extract and subsequent analysis of the RNA products were identical to those described above. The probe used for the S1-nuclease protection analysis was made from *HindIII*-linearized and kinase-labelled major late promoter-containing plasmid. After secondary digestion with *SmaI*, the probe was made single-stranded and isolated by the method of Maxam and Gilbert³³. S1-protection analysis was performed as described above, except that hybridization was at 46 °C.

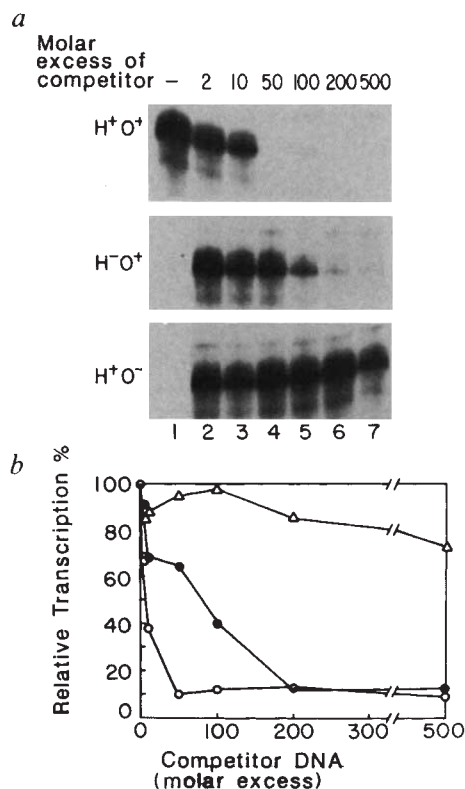


Fig. 4 Relative sensitivity of transcription from the H^+O^+ wild-type promoter to wild-type (H^+O^+) or heptamer- (H^-O^+) or octamer-mutated (H^+O^-) competitor oligonucleotides. Transcription reactions were carried out in Namalwa nuclear extract in the absence (lane 1) or presence (lanes 2–7) of increasing concentrations of competitor DNA. *a*, S1-nuclease protection analysis of the transcription products. *b*, Quantitation of transcription in the presence of the indicated concentrations of H^+O^+ (open circles), H^-O^+ (closed circles) or H^+O^- (triangles) competitor DNA. The level of transcription is shown relative to transcription in the absence of competitor. Transcription was in the presence of 2.4 mg nuclear extract per ml and 10 μ g template per ml and the product analysed as described in Fig. 3. The S1 nuclease-protected products were excised and quantitated by scintillation counting.

12. Poellinger, L. & Roeder, R. G. *Molec. cell. Biol.* **9**, 747–756 (1989).
13. Fletcher, C., Heintz, N. & Roeder, R. G. *Cell* **51**, 773–781 (1987).
14. O'Neill, E. A. *et al. Science* **241**, 1210–1213 (1988).
15. Scheidereit, C., Heguy, A. & Roeder, R. G. *Cell* **51**, 783–793 (1987).
16. Scheidereit, C. *et al. Nature* **336**, 551–557 (1988).
17. Landolfi, N. F., Capra, D. J. & Tucker, P. W. *Nature* **323**, 548–551 (1986).
18. Staudt, L. M. *et al. Nature* **323**, 640–643 (1986).
19. Ko, H.-S., Fast, P., McBride, W. & Staudt, L. M. *Cell* **55**, 135–144 (1988).
20. Weinberger, J., Baltimore, D. & Sharp, P. A. *Nature* **322**, 846–848 (1986).
21. Westin, G., Gerster, T., Muller, M., Schaffner, G. & Schaffner, W. *Nucleic Acids Res.* **15**, 6787–6798 (1987).
22. LeBowitz, J. H., Kobayashi, T., Staudt, L., Baltimore, D. & Sharp, P. A. *Genes Dev.* **2**, 1227–1237 (1988).
23. Hochschild, A. & Ptashne, M. *Cell* **44**, 681–687 (1986).
24. Robertson, M. *Nature* **327**, 464–466 (1987).
25. Ptashne, M. *Nature* **335**, 683–689 (1988).
26. Horikoshi, M., Carey, M. F., Kakidani, H. & Roeder, R. G. *Cell* **54**, 665–669 (1988).
27. VanDyke, M., Roeder, R. G. & Sawadogo, M. *Science* **241**, 1335–1338 (1988).
28. Horikoshi, M., Hai, T., Lin, Y.-S., Green, M. & Roeder, R. G. *Cell* **54**, 1033–1042 (1988).
29. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. *Science* **240**, 1759–1764 (1988).
30. Robertson, M. *Nature* **336**, 522–524 (1988).
31. Dignam, J. D., Lebowitz, R. M. & Roeder, R. G. *Nucleic Acids Res.* **11**, 1475–1489 (1983).
32. Weaver, R. F. & Weissman, C. *Nucleic Acids Res.* **7**, 1175–1193 (1979).
33. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).

Crystal structure of a retroviral protease proves relationship to aspartic protease family

Maria Miller*, Mariusz Jaskólski†, J. K. Mohana Rao*, Jonathan Leis† & Alexander Wlodawer*

* Crystallography Laboratory, NCI-Frederick Cancer Research Facility, BRI-Basic Research Program, PO Box B, Frederick, Maryland 21701, USA

† Department of Biochemistry, Case Western Reserve University, Cleveland, Ohio 44106, USA

Retroviral gag, pol and env gene products are translated as precursor polypeptides, which are cleaved by virus-encoded proteases to produce the mature proteins found in virions^{1–11}. On the basis of the conserved Asp–Thr/Ser–Gly sequence at the putative protease active sites, and other biochemical evidence^{2,3,12–16}, retroviral proteases have been predicted to be in the family of pepsin-like aspartic proteases. It has been suggested that aspartic proteases evolved from a smaller, dimeric ancestral protein¹⁷, and a recent model of the human immunodeficiency virus (HIV) protease postulated that a symmetric dimer of this enzyme is equivalent to a pepsin-like aspartic protease¹⁸. We have now determined the crystal structure of Rous sarcoma virus (RSV) protease at 3-Å resolution and find it is dimeric and has a structure similar to aspartic proteases^{19–22}. This structure should provide a useful basis for the modelling of the structures of other retroviral proteases, such as that of HIV, and also for the rational design of protease inhibitors as potential antiviral drugs.

Trigonal crystals containing two molecules of the protease in the asymmetric unit²³ were grown under conditions in which the protease remains active¹². The crystallographic model of the RSV protease (PR) was obtained as described in the Fig. 1 legend and is shown in Fig. 2. Only one molecule was built independently. It was then rotated around the non-crystallographic dyad to yield the RSV PR dimer. The general topology of the molecule can be described using the conventions applied to a domain of an aspartic protease²¹. The N-terminal α β -strand (residues 1–7) is a part of the dimer interface, whereas in aspartic proteases it belongs to the inter-domain junction. The β β -strand (8–21) continues through a broad loop (22–27) into the ϵ β -chain which terminates at the active-site triplet (37–39). In aspartic proteases, the d -chain is followed by the $h\alpha$ -helix. In the retroviral molecule this helix is limited to one turn (46–49) but is immediately followed by a very characteristic amino-acid

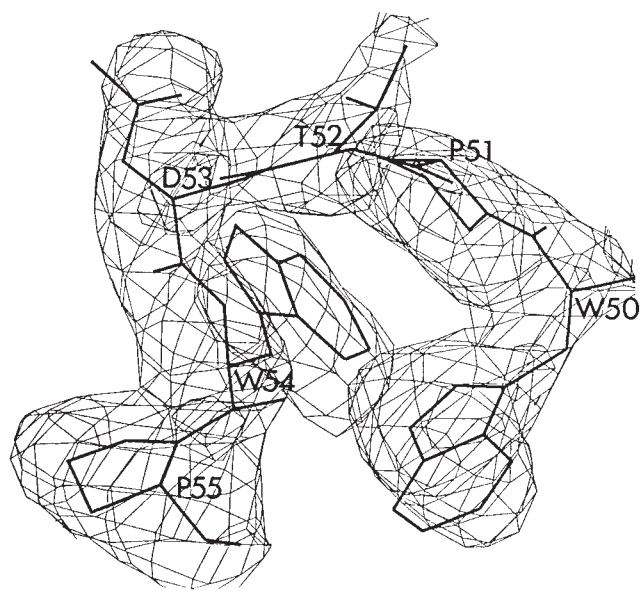
promoter activity sensitive to subtle fluctuations in cellular OTF levels and thereby modulate promoter activity during, for instance, lymphoid development. Given the strong OTF-2-dependent transcription signal *in vitro*, the H^+O^+ minimal promoter could be used to characterize functional domains of OTF-2 and to investigate the communication of OTF-2 with the TATA factor TFIID and RNA polymerase II in the propagation of a transcription signal.

We thank Dr Thomas Gerster for discussion. This work was supported by the National Cancer Institute and the Pew Trusts (R.G.R.), the Cancer Research Institute and the National Institutes of Health (B.K.Y.) and by the Swedish Medical Research Council (L.P.).

Received 2 November; accepted 21 December 1988.

1. Falkner, F. G. & Zachau, H. G. *Nature* **310**, 71–74 (1984).
2. Parslow, T. G., Blair, D. L., Murphy, W. J. & Granner, D. K. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2650–2654 (1984).
3. Siu, G., Springer, E. A., Huang, H. V., Hood, L. E. & Crews, S. T. *J. Immun.* **138**, 4466–4471 (1987).
4. Eaton, S. & Calame, K. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7634–7638 (1987).
5. Mason, J. O., Williams, G. T. & Neuberger, M. S. *Cell* **41**, 479–487 (1985).
6. Ballard, D. W. & Bothwell, A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9626–9630 (1986).
7. Mizushima-Sugano, J. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8511–8515 (1986).
8. Dreyfus, M., Doyen, N. & Rougeon, F. *EMBO J.* **6**, 1685–1690 (1987).
9. Wirth, T., Staudt, L. & Baltimore, D. *Nature* **329**, 174–178 (1987).
10. Gerster, T., Matthias, P., Thali, M., Jiricny, J. & Schaffner, W. *EMBO J.* **6**, 1323–1330 (1987).
11. Lenardo, M., Pierce, J. W. & Baltimore, D. *Science* **236**, 1573–1577 (1987).

† Permanent address: A. Mickiewicz University, Faculty of Chemistry, Poznań, Poland.



sequence loop, WPTDWP (in the one-letter code) (50-55) which could be considered a distorted continuation of *h* (Fig. 1). The rest of the molecule (except the C-terminal chain) has a topology related to that described above by an approximate intramolecular twofold symmetry (Fig. 2*a*). Residues 56-61 form the α' -chain and, as in pepsin-like proteases, lead to the 'flap' loop. The h' -helix (110-118) is followed by a straight C-terminal β -chain (119-124) which also belongs to the dimer interface and can be designated *q*. In contrast, pepsin-like proteases have a double-stranded (*q* and *r*) β -sheet at this position in the inter-domain interface. Even though the detailed connectivity in the interface region is different in these two types of proteases, the overall topology is similar and can be described as an antiparallel β -sheet consisting of six strands in pepsin-like proteases or four in the retroviral protease (Fig. 2*b*). This sheet forms the bottom of the dimer on which the active-site loops are situated. Two 'flap' arms (one from each monomer) project over the active site in the retroviral protease whereas in pepsin-like proteases only one prominent flap (in the N-terminal domain) has been described. There is a large cleft between the active site and the flaps which can be recognized as the substrate-

Fig. 1 Electron density and model in the loop containing residues 50-55 of the retroviral protease.

Methods. The protein used in this investigation was purified from RSV Pr-C strain, and was crystallized as previously described²³. The crystals belong to the trigonal space group $P3_121$ with unit cell parameters $a = 88.95$ Å, $c = 78.90$ Å and with two molecules in the asymmetric unit. Diffraction data for the native crystals and several heavy atom derivatives were taken with an electronic area detector²⁴. The space group enantiomorph was determined from anomalous scattering data. The structure was solved by multiple isomorphous replacement method²⁵ as implemented in the PROTEIN program package²⁶. A difference Patterson map for an uranyl acetate derivative revealed a single, highly occupied site. Three mercury derivatives were solved using difference Fourier and Patterson maps. A combined lack-of-closure refinement of the four derivatives, including anomalous scattering for the uranyl derivative, gave a figure of merit of 0.67 and the final parameters listed in Table 1. The map shown here was calculated using MIR phases improved by solvent flattening²⁷ and, after the two molecules in the asymmetric unit were identified, by symmetry averaging (map correlation and averaging programs written by J.K.M.R., unpublished). The electron density corresponding to the amino-acid sequence WPTDWP was recognized in the map and served as a starting point for sequence fitting using the program system FRODO²⁸. With the exception of eight amino acids in the 'flap' region, all other residues could be identified in the map. Several cycles of restrained least-squares refinement²⁹ converged at $R = 0.29$ for 10-3 Å data.

binding pocket. Located on both sides of the interface region are the central cores of the two monomers. A characteristic topological feature in these cores is a sandwich of two four-stranded β -sheets with perpendicular chain directions (Fig. 2). One of these sheets has antiparallel β -chains: *c*(28-34), *b*(13-21), *b'*(80-89), *c'*(92-98), whereas the other sheet can be described as being composed of two antiparallel Greek letters ψ superimposed in such a way that the central bar in each letter makes an outer bar in the other. Only the central chains in this sheet are antiparallel; the outer ones are parallel to their neighbours. One of the ψ letters comprises chains *c*(35-39), *d*, *d'* and the other is made up of *c'*(99-104), *d'*, *d*.

The active site of the RSV protease, sequence Asp37-Ser38-Gly39 on each chain, is typical for an aspartic protease. The residue immediately following the triplet is always alanine in retroviral proteases (Fig. 3*a*) whereas in all pepsin-like enzymes (except in one domain of human renin where alanine follows the triplet) this residue has a hydroxyl group which usually hydrogen-bonds to the catalytic Asp side chain on the same strand (Fig. 3*b*).

The amino-acid configuration around the active site in the

Table 1 Heavy-atom parameters after lack-of-closure refinement at 3 Å

Derivative	Site	Occ*	x	y	z	B (Å ²)	R _C †
Uranyl acetate	1	2.8	-0.080	0.457	-0.011	31.5	0.49
Mercury acetate	1	1.3	0.498	0.460	0.127	36.2	0.61
	2	0.7	0.755	0.456	0.072	32.7	
	3	1.1	-0.082	0.457	-0.013	26.1	
	4	0.5	0.798	0.516	0.034	36.3	
Mersalyl	1	1.2	0.498	0.460	0.127	30.5	0.69
	2	1.2	0.753	0.456	0.071	32.8	
	3	0.5	-0.080	0.457	-0.013	16.6	
	4	0.5	0.571	0.079	0.084	38.8	
Methyl mercury acetate (4 Å)	1	0.8	0.502	0.462	0.127	26.7	0.68
	2	1.1	0.756	0.457	0.071	26.3	
	3	0.6	0.472	0.495	0.152	37.8	

The top three derivatives were obtained by soaking in 2 mM solutions of heavy atoms for 24-48 h; a six-day soak in 6 mM methyl mercury acetate produced the fourth. Uranyl ion binds between the two aspartates in the active site. Sites 1 and 2 of all mercurials are due to binding to Cys113 in the two molecules in the asymmetric unit. Site 3 of mercury acetate and mersalyl is found in the active site.

* Occupancy in arbitrary units.

† Cullis R-factor = $(\sum |F_{h(obs)} - F_{h(calc)}|) / \sum F_{h(obs)}$.



Fig. 2 Chain tracing of RSV PR. *a*, Stereo view of the C α chain of a single molecule. Residues 63–70, part of the flap region, are not visible in the available map and may be disordered in the crystal. The fragment of the flap region that has been traced so far is long enough to extend over the active site. The molecule has an approximate twofold symmetry, the corresponding local axis being horizontal in the plane of the figure. The r.m.s. deviation for 40 pairs of C α atoms related by this intramolecular axis is 2.6 Å. (This, and all the other alignments described here used an unpublished computer program written by J.K.M.R.) Similar observations have been previously reported for pepsin³⁰. *b*, Stereoview of the RSV PR dimer along the non-crystallographic dyad. The dimer interface is a four-stranded antiparallel β -sheet. Two additional outer strands are in extended configuration and do not form proper β -hydrogen bonds beyond the initial turn.

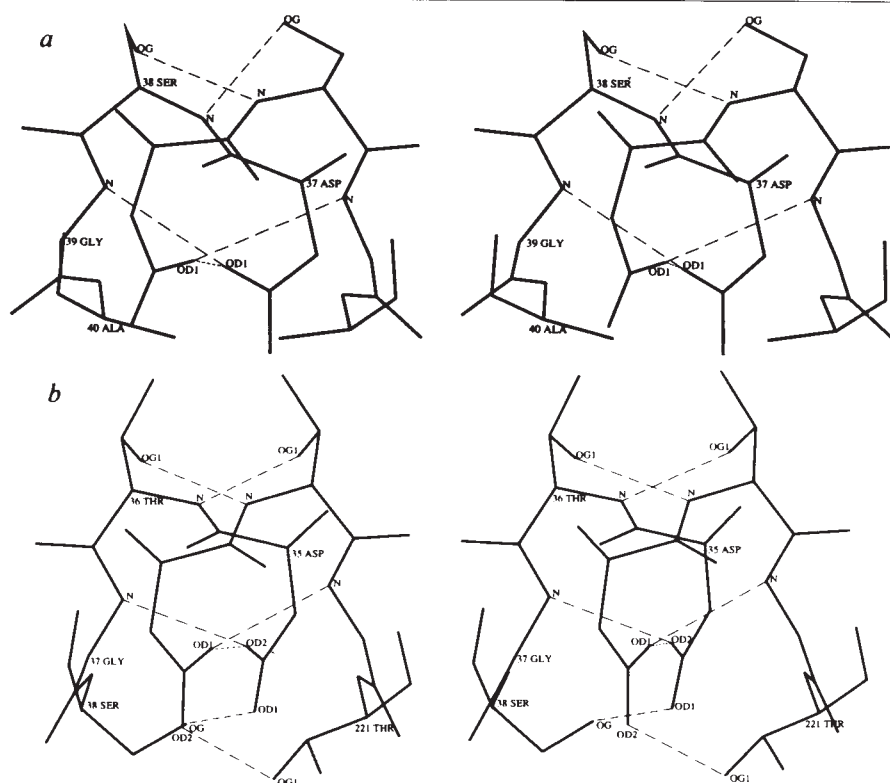


Fig. 3 Stereoviews of the active sites of aspartic proteases. *a*, Retroviral protease, showing the hydrogen bonding of the 'fireman's grip'²¹ through the serines, and an interaction between the two Asp side chains. *b*, Rhizopuspepsin²⁰, with additional interactions of the hydroxyls at Ser38 and Thr221 with Asp35 and Asp218, respectively.

RSV PR aligns well with the analogous region in rhizopuspepsin²⁰. Using the C α atoms, residues 30–49 or RSV PR (including the active site triplet) can be superimposed over the stretch 28–47 in the N-terminal domain of rhizopuspepsin with an r.m.s. deviation of 2.6 Å. An even longer stretch (26–55) in RSV PR aligns well with the region 207–236 in the C-terminal domain of rhizopuspepsin with an r.m.s. deviation of 3.1 Å. The region 97–124 in RSV PR that possesses the conserved sequence in the h' helix with an invariant Gly residue can be aligned with the 287–314 stretch at the C-terminal end of rhizopuspepsin to an r.m.s. deviation of 2.8 Å.

The RSV PR molecule described above has been compared with a slightly modified variant (L. Pearl, personal communication) of the model of the HIV protease¹⁸. HIV PR consists of only 99 amino acids but 25 of these are identical to those in the 124 amino-acid RSV PR when the two proteases are structurally aligned. The alignment was obtained by superimposing the active-site triplet and a second conserved sequence, Gly-Arg-Asp, at the start of the h'-helix in RSV PR on their corresponding sequences in HIV PR and by permitting the deletions in the HIV PR sequence to occur only at surface turns and loops of RSV PR. Superposition of the HIV PR model on the RSV PR molecule yields an r.m.s. deviation of 2 Å for 36 C α atoms. The superimposed segments include strands c (involving the active site), d and d' and the h' helix. Although the discrepancies are larger in other regions, the overall topology proposed for HIV PR almost exactly matches that of the RSV PR, with the a-b and a'-b' (flap) loops following parallel courses with an offset of 3–4 Å. The only major discrepancy is at the C terminus where intermolecular contacts are used to produce the dimer in the present structure. Most of the residues conserved between the two proteases are on the intermolecular interface. A new model of the HIV PR has now been constructed by us on the basis of the RSV PR structure. Since processing of retrovirus protein is essential for production of infectious virus particles^{2,4,5,7,8} and the avian sarcoma and leukemia viruses protease (ASLV PR) can process murine leukaemia virus (MuLV)^{9,10}, feline leukaemia virus (FeLV)¹¹ GAG polyprotein precursors, and HIV peptide substrates⁴, the structure defined in this report should be valuable in designing inhibitors directed towards the HIV PR.

We would like to thank Dr I. Weber for modelling the HIV PR, Drs O. Herzberg, D. Davies, A. Skalka and J. Jentoft for discussions and encouragement, and the Center for Advanced Research in Biotechnology for access to their data collection facility. This research was sponsored in part by the National Cancer Institute, Department of Health and Human Services (DHHS) with Bionetics Research, Inc. and in part by the NIH. The contents of this publication do not necessarily reflect the views or policies of the DHHS, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government.

Note added in proof: The structure discussed above has now been refined at 2-Å resolution, resulting in only small adjustments of the model and in the crystallographic R-factor of 0.16.

15. Kotler, M., Katz, R., Danho, W., Leis, J. & Skalka, A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4185–4189 (1988).
16. Schneider, J. & Kent, S. *Cell* **54**, 363–368 (1988).
17. Tang, J., James, M. N. G., Hsu, I. N., Jenkins, J. A. & Blundell, T. L. *Nature* **271**, 618–621 (1978).
18. Pearl, L. & Taylor, W. *Nature* **329**, 351–354 (1987).
19. Andreeva, N. S., Zdanov, A. S., Gustchina, A. E. & Fedorov, A. A. *J. biol. Chem.* **259**, 11353–11365 (1984).
20. Suguna, K. *et al. J. molec. Biol.* **196**, 877–900 (1987).
21. Blundell, T., Jenkins, J., Pearl, L., Sewell, T. & Pedersen, V. in *Aspartic Proteinases and their Inhibitors* (ed. Kostka, V.) 151–161 (Walter de Gruyter, Berlin, 1985).
22. James, M. N. G. & Sielecki, A. R. *J. molec. Biol.* **163**, 299–361 (1983).
23. Miller, M., Leis, J. & Wlodawer, A. *J. molec. Biol.* **204**, 211–212 (1988).
24. Howard, A. J. *et al. J. appl. Crystallogr.* **20**, 383–387 (1987).
25. Blundell, T. L. & Johnson, L. N. *Protein Crystallography* (Academic, London, 1976).
26. Steigemann, W. thesis, Technical Univ. Munich (1972).
27. Wang, B. C. in *Methods in Enzymology* Vol. 115 (eds Wyckoff, H. W., Hirs, C. H. & Timasheff, S. N.) 90–112 (Academic, London, 1985).
28. Jones, T. A. *J. appl. Crystallogr.* **11**, 268–272 (1978).
29. Hendrickson, W. A. in *Methods in Enzymology* Vol. 115 (eds Wyckoff, H. W., Hirs, C. H. & Timasheff, S. N.) 252–270 (Academic, London, 1985).
30. Andreeva, N. S. *et al. Molec. Biol. Engl. Trans. Molec. Biol. (Mosc)* **12**, 704–716 (1978).

Dynamics of the multidomain fibrinolytic protein urokinase from two-dimensional NMR

Robert E. Oswald*†, Michael J. Bogusky*, Michelle Bamberger*†, Richard A. G. Smith‡ & Christopher M. Dobson*§

* Inorganic Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QR, UK

‡ Beecham Pharmaceuticals Research Division, Biosciences Research Centre, Great Burgh, Yew Tree Bottom Road, Epsom, Surrey KT18 5XQ, UK

The recent demonstrations that thrombolytic therapy with plasminogen activators can result in substantial reductions in mortality from coronary thrombosis^{1–3} have generated considerable interest in the properties of fibrinolytic enzymes. Examination of the primary sequence of these proteins⁴ (which include tissue plasminogen activator, plasminogen, and urokinase) reveals that each is composed of a mosaic of domains which appear to be spatially distinct and connected by short peptide linkers. There is, however, little experimental information about the three-dimensional structure of any of the proteins, although several X-ray diffraction and NMR studies of isolated domains have been reported^{5–10}. Here we report two-dimensional NMR spectra of intact urokinase which are remarkably well resolved for a protein of this molecular weight. This effect is a consequence of substantial independent motion between individual domains of the protein, which overcomes the broadening effects anticipated for the slow overall tumbling rate of the intact molecule. As well as having significance for the physiological role of the protein, these results provide a direct means for the comparison of structural features determined for the isolated domains with those of the intact protein and may provide a basis for proposing or evaluating models for the overall structure of fibrinolytic proteins^{11,12}. Preliminary results with other proteins indicate that this approach may be generally applicable to other multidomain proteins of the fibrinolytic family.

The complete two-dimensional (2-D) NOESY (above diagonal) and COSY (below diagonal) spectra of urokinase in D₂O are shown in Fig. 1. Urokinase is comprised of two structural domains, an epidermal growth factor (EGF)-like domain and a kringle domain, in addition to a glycosylated serine protease domain, resulting in an overall relative molecular mass of ~54,000. The spectral resolution obtained is remarkable for a macromolecule of this size. Despite the complexity of the spectrum, numerous well resolved crosspeaks are present in

1. Dickson, C., Eisenman, R., Fan, H., Hunter, E. & Teich, N. in *RNA Tumor Viruses* Vol. 1 (eds Weiss, R., Teich, N., Varmus, H. & Coffin, J.) 513–648 (Cold Spring Harbor Laboratory, New York, 1984).
2. Katoh, I., Yasunaga, T., Ikawa, Y. & Yoshinaka, Y. *Nature* **329**, 654–656 (1987).
3. Seelmeier, S., Schmidt, H., Turk, V. & von der Helm, K. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6612–6616 (1988).
4. Voynow, S. & Coffin, J. *J. Virol.* **55**, 79–85 (1985).
5. Katoh, I. *et al. Virology* **145**, 280–292 (1985).
6. Alexander, F. *et al. J. Virol.* **61**, 534–542 (1987).
7. Kohl, N. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 4686–4690 (1988).
8. Felsenstein, K. & Goff, S. *J. Virol.* **62**, 2179–2182 (1988).
9. Demsey, A., Collins, F. & Kawka, B. *J. Virol.* **36**, 872–876 (1980).
10. Yoshinaka, Y. & Luftig, R. *Virology* **111**, 239–250 (1981).
11. Khan, A. & Stephenson, J. *J. Virol.* **29**, 649–656 (1979).
12. Kotler, M., Danho, W., Katz, R. A., Leis, J. & Skalka, A. M. *J. biol. Chem.* (in the press).
13. Toh, H., Ono, M., Saigo, K. & Miyata, T. *Nature* **315**, 691–692 (1985).
14. Miller, R. *Science* **236**, 722–725 (1987).

† Permanent address: Department of Pharmacology, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA.

§ To whom correspondence should be addressed.

Received 2 November 1988; accepted 6 January 1989.

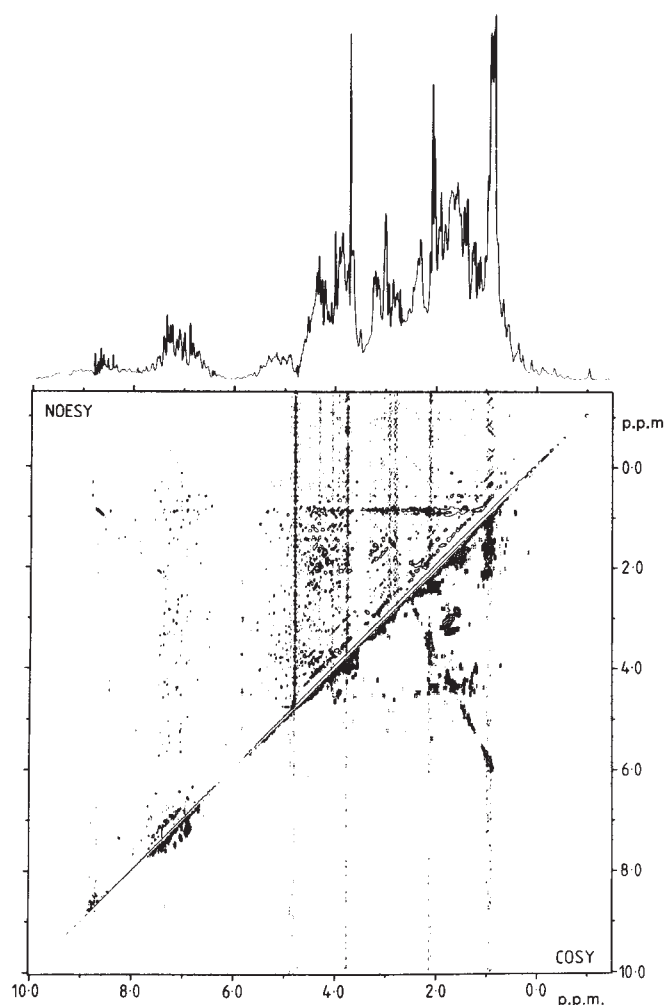


Fig. 1 600 MHz phase-sensitive 2-D ^1H -NMR spectra of active site-blocked urokinase (1.5 mM in D_2O , pH 4.5, 29 $^\circ\text{C}$), prepared as described elsewhere¹⁴. The portion of the plot above the diagonal is a NOESY (nuclear Overhauser enhancement spectroscopy) spectrum. The spectrum was acquired with 2,048 points in the t_2 dimension and 256 increments in the t_1 dimension using a mixing time of 150 ms (random modulation of the mixing period by $\pm 5\%$). Crosspeaks in a NOESY spectrum represent dipolar couplings between protons which are close in space (≤ 5 Å). Crosspeaks between aromatic and aliphatic resonances and between amide and aliphatic resonances are, in general, indicative of residues in structured regions of the protein which are in close proximity. The portion of the spectrum below the diagonal corresponds to the double quantum filtered 2-D correlated (DQF-COSY) spectrum. Crosspeaks in a COSY spectrum correlate proton resonances which are scalar coupled and produce characteristic patterns for specific amino-acid spin systems. Both positive and negative contours are plotted. All spectra were acquired at 600.13 MHz on a Bruker AM-600 spectrometer with spectral widths of 8,064 kHz in both dimensions. The time-domain data matrix was multiplied by a 20° phase-shifted sine bell in both dimensions and zero filled to 1,024 in t_1 before Fourier transformation, resulting in a 1K by 1K frequency matrix. The data were processed on a MicroVAX II computer using software provided by Dr D. R. Hare (FTNMR; Hare Research, Inc.) with minor modifications. The baseline was corrected in the t_2 dimension by a cubic spline interpolation.

both the NOESY and COSY spectra. Some of these resonances may arise in part from unstructured regions of the protein which exhibit rapid motion superimposed on the overall tumbling of the molecule. This can result in the narrowing of resonance lines even of large proteins¹³, to the extent that crosspeaks are anticipated in the 2-D spectra. Close inspection of Fig. 1, however, reveals that many of the crosspeaks in the NOESY spectrum of urokinase must arise from residues in highly structured regions of the protein. For example, resonances in the region between 5 and 6 p.p.m., characteristic of C_α protons in β -sheet secondary structure, show intense NOE crosspeaks with aromatic (6.5–8.5 p.p.m.), C_α (5–6 p.p.m.) and C_β (1–4 p.p.m.) protons. Additionally, numerous NOE crosspeaks are apparent between amide (6.5–10 p.p.m.) and aliphatic (0–5 p.p.m.) proton resonances. These effects arise from the close spatial proximity (< 5 Å) of the protons and are indicative of a compact, globular structure. The COSY spectrum in Fig. 1 reveals the presence of a large number of correlations between scalar coupled protons. Typically, for a protein of this molecular weight, cancellation of the antiphase components of the COSY crosspeaks would be expected as the linewidths approach or exceed the ^1H - ^1H scalar coupling, resulting in little or no crosspeak intensity; this is clearly not the case for urokinase.

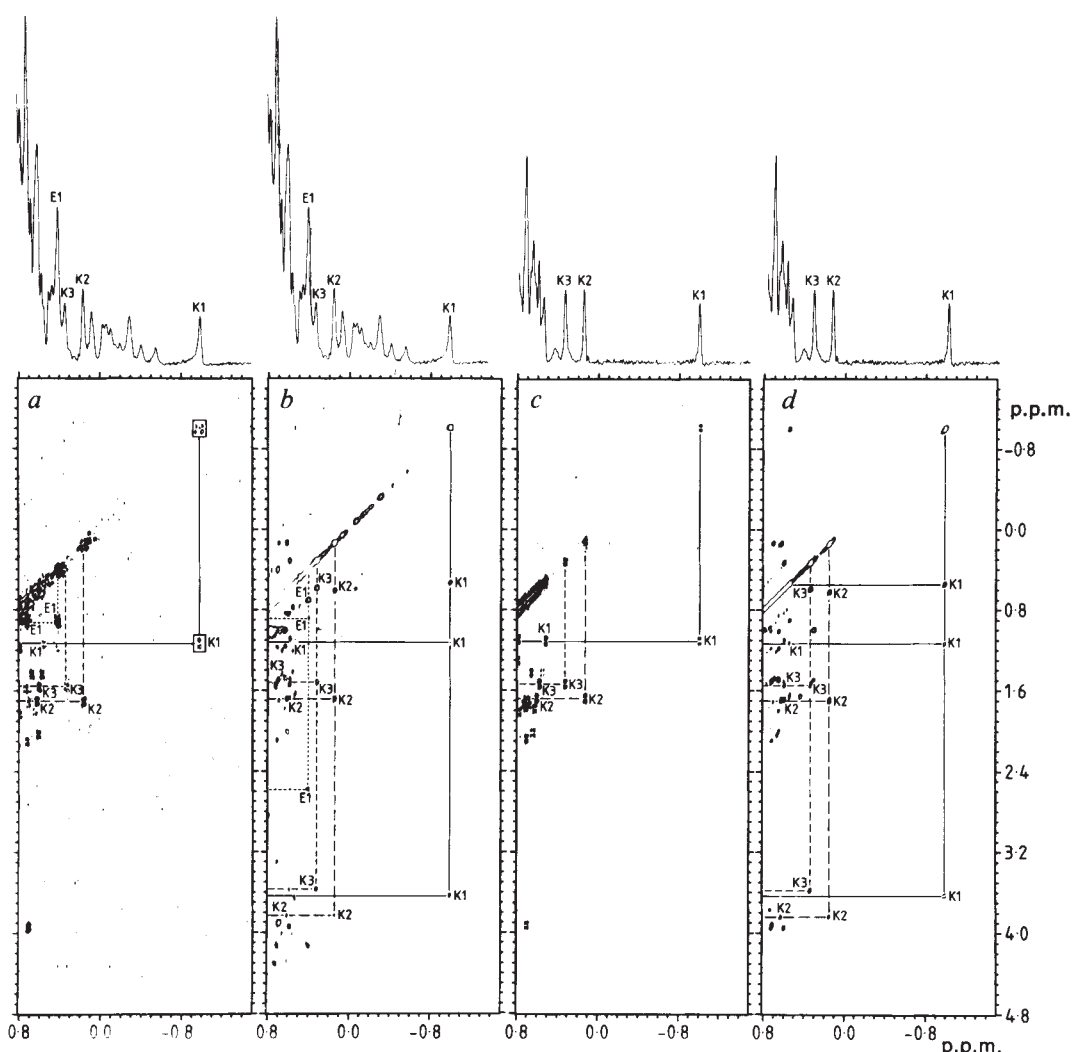
Comparison of the upfield region of the 1-D spectra of intact urokinase (Fig. 2) with the spectra of the proteolytically derived serine protease, kringle and EGF-kringle domains indicated that the majority of the resonances arise from the serine protease domain, in addition to three resonances from the kringle domain and one from the EGF domain¹⁴. The 2-D spectra of intact urokinase are, however, remarkably similar to those of the

kringle domain (with the exception of several crosspeaks arising from the EGF domain), indicating that a significant portion of the crosspeak intensity in the spectra of urokinase must arise from the kringle domain. Not only are the crosspeaks attributed to the kringle domain sufficiently well resolved to allow identification of the amino-acid residue type, but they are present in virtually identical positions in both spectra, indicating close structural similarities between the excised fragment and the domain in the intact protein. Similar results are evident in the aromatic and other regions of the spectra.

These results indicate that the linewidths of resonances attributed to the protease domain must be sufficiently broad to result in cancellation or significant reduction of crosspeak intensity, whereas resonances from the kringle and EGF domains are considerably narrower and are, therefore, clearly observable in the 2-D spectra. The 2-D NMR experiments, in effect, enhance the crosspeak intensities of the narrow resonances (those with longer T_2) relative to the broader resonances¹⁵. This results in the selective observation of resonances from the smaller domains of the protein. We propose that the narrower linewidths must be a result of substantial motion between the EGF-kringle domains and the protease domain, superimposed on the overall tumbling of the molecule, with the smaller domains reorienting faster than the larger domain.

Under appropriate conditions (pH 4.5, 59 $^\circ\text{C}$), an equilibrium is established between the folded and unfolded states of the EGF-kringle domains in urokinase (full details are reported elsewhere¹⁴). As a consequence of the rate of interconversion between the two states, substantial broadening of EGF-kringle resonances occurs as a result of chemical exchange effects, and

Fig. 2 Comparison of the upfield region of the 600 MHz phase-sensitive 2-D¹H-NMR spectra of urokinase and its proteolytically-derived fragment (Ser 47-Lys 135) which corresponds to the kringle domain. This region of the spectrum includes resonances of methyl groups which are in close proximity to aromatic residues. The kringle domain was produced as described elsewhere¹⁴. All spectra were obtained at 29 °C using sample concentrations of 1.5 mM at pH 4.5. Included are sections of: *a*, a double quantum filtered COSY spectrum of urokinase, *b*, a NOESY spectrum of urokinase, *c*, a double quantum filtered COSY spectrum of the kringle domain, and *d*, a NOESY spectrum of the kringle domain. Several amino-acid spin systems are illustrated and identified as belonging to the kringle (K1, K2, K3) or EGF (E1) domains. K1 represents a leucine residue and has tentatively been assigned to Leu 96 (urokinase numbering²²; equivalent to Leu 45 with kringle numbering) in the kringle domain by analogy with studies performed on kringles 4 and 5 isolated from human plasminogen⁵⁻⁸. This residue is conserved in all known kringle sequences. Spin systems K2 and K3 have been identified as valines in both the kringle domain and the intact urokinase. On the basis of a comparison of 1D ¹H-NMR spectra of kringle and EGF-kringle fragments, the spin system E1 is derived from the EGF domain, but the nature of its spin system has yet to be identified. Boxed regions were plotted at a lower contour level.



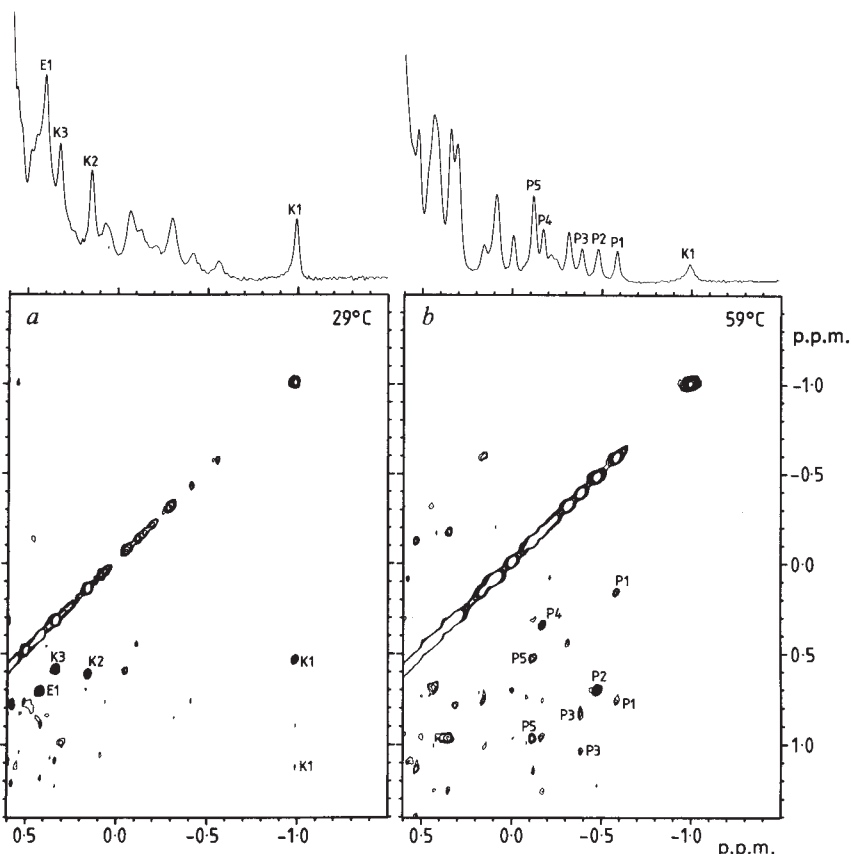
narrower linewidths can be observed for protease resonances arising from the increased rate of tumbling of the molecule. Comparison of the 2-D NOESY spectra of urokinase at 29 °C (Fig. 3a) and 59 °C (Fig. 3b) reveals that at 29 °C crosspeak intensity is only apparent from residues in the EGF and kringle domains, but at 59 °C resonances from the protease domain are selectively enhanced. Alternatively, a comparison of spectra above and below the unfolding transition can provide similar information, even in cases where exchange broadening is not detectable. Recent results on the isolated EGF-kringle fragment suggest that conditions exist whereby the kringle domain can be independently unfolded relative to the EGF and protease domains. This would allow, in principle, extension of these approaches to provide selective observation of many of the resonances in structured regions of the three domains. Detailed comparisons could then be made between the structures of domains in isolated fragments and in the intact protein, which would allow interactions between the individual domains to be studied.

These results indicate that the EGF and kringle domains of urokinase behave as independent structural and dynamic entities relative to the protease domain. Such dynamic independence of the domains may be common to other members of this protein

family. Indeed, we have found in preliminary experiments that, as with urokinase, well resolved resonances in ¹H NMR spectra can be observed from individual domains of plasminogen and tissue plasminogen activator. This is also consistent with studies of the reversible thermal unfolding of these proteins (refs 14 and 16; M. J. B. *et al.*, unpublished results) and may explain, in part, the difficulty in crystallizing proteins of this type. The degree of domain independence will, however, presumably be a function of the nature of interactions between the different domains, as well as the length of the linker region connecting the various domains. It is interesting in this regard that the linker region between the kringle and serine protease domains in urokinase (residues 132 to 147) is one of the longer linker sequences found in the family of fibrinolytic proteins⁴.

Although the results presented here for urokinase were obtained at pH 4.5, similar spectra were obtained at pH 7.5, indicating the presence of independent domain motion under physiological conditions. Functional independence of the domains of urokinase may be of physiological importance because the amino-terminal domains seem to anchor the serine protease domain to the cell surface through a specific receptor¹⁷. The results presented here indicate that the serine protease domain could remain structurally independent in those circum-

Fig. 3 Comparison of the upfield region of the NOESY spectrum of urokinase at *a*, 29 °C and at *b*, 59 °C. At the higher temperature, the EGF and kringle domains are partially unfolded under these conditions¹⁴. The resolved resonances from the EGF and kringle domains (E1, K1, K2, K3) are absent from the spectrum as a consequence of substantial broadening resulting from chemical exchange. This is readily apparent in the 1-D spectrum at the top of the figure. In contrast to the spectrum recorded at 29 °C, the resonances attributed to the serine protease domain (labelled P1–P5) are significantly narrowed at the higher temperature, revealing pronounced crosspeaks between the protease resonances. The narrowing of the protease resonances occurs as a result of the increased rotational correlation time of the intact protein at higher temperature. This figure clearly demonstrates the selective observation of resonances from either the EGF-kringle or protease domains as a function of experimental conditions.



stances because of flexibility in the linker region joining it to the EGF-kringle domain. This could leave it free to participate in cell-associated control of tissue remodelling and metastasis. NMR studies of the type described here for urokinase, and in preliminary form for plasminogen and tissue plasminogen activator, could be crucial in providing an experimental basis for understanding the structural and dynamic properties of the fibrinolytic proteins. These properties will be important to the design of mutant and hybrid plasminogen activators in which domain linker construction has hitherto been based mainly on genetic considerations^{18–21}.

We thank Andrew J. Teuten and Li Xiang for valuable discussions. This work is supported in part by the Science and Engineering Research Council. R.E.O. is supported by a senior fellowship from the National Institutes of Health and a Fulbright Fellowship. M.J.B. is supported by a National Science Foundation postdoctoral fellowship. C.M.D. is a member of the Oxford Centre for Molecular Sciences.

Received 4 November; accepted 22 December 1988.

- GISSI Study Group *Lancet* ii, 871–874 (1987).
- Aims Trial Study Group *Lancet* i, 545–549 (1988).
- Wilcox, R. G. *et al.* *Lancet* ii, 525–530 (1988).
- Bachmann, F. & Kruithof, E. K. O. *Sem. Thromb. Hemostas.* 10, 6–17 (1984).
- Williams, R. J. P., Esnov, P., Lawrence, M. & Cederholm-Williams, F. A. *FEBS Lett.* 209, 111–116 (1986).
- Ramesh, V., Petros, A. M., Llinás, M., Tulinsky, A. & Park, C. H. *J. molec. Biol.* 198, 481–498 (1987).
- Thewes, T., Ramesh, V., Simplaceanu, E. & Llinás, M. *Eur. J. Biochem.* 175, 237–249 (1988).
- Mabbutt, B. C. & Williams, R. J. P. *Eur. J. Biochem.* 170, 539–548 (1988).
- Park, C. H. & Tulinsky, A. *Biochemistry* 25, 3977–3982 (1986).
- Harlos, K., Boys, C. W. G., Holland, S. K., Esnouf, M. P. & Blake, C. C. F. *FEBS Lett.* 224, 97–103 (1987).
- Holland, S. K. & Blake, C. C. F. *EMBO J.* (in the press).
- Tranqui, L., Prandini, M. H. & Chapel, A. *Biol. cellulaire* 34, 39–42 (1979).
- Perham, R. N., Duckworth, H. W. & Roberts, A. C. K. *Nature* 292, 474–477 (1981).
- Bogusky, M. J., Dobson, C. M. & Smith, R. A. G. *Biochemistry* (submitted).
- Weiss, M. A., Eliason, J. L. & States, D. J. *Proc. natn. Acad. Sci. U.S.A.* 81, 6019–6023 (1984).
- Novokhasny, V. V., Kvainov, S. A. & Privalov, P. L. *J. molec. Biol.* 179, 215–232 (1984).
- Blasi, F. *Fibrinolysis* 2, 73–84 (1988).
- Harris, T. J. R. *Protein Eng.* 1, 449–458 (1987).
- Nelles, L., Lijnen, H. R., Collen, D. & Holmes, W. E. *J. biol. Chem.* 262, 10855–10862 (1987).
- Gheysen, D. *et al.* *J. biol. Chem.* 262, 11779–11784 (1987).
- de Vries, C., Veerman, H., Blasi, F. & Pannekoek, H. *Biochemistry* 27, 2561–2572 (1988).
- Holmes, W. E. *et al.* *Biotechnology* 3, 923–929 (1985).

GUIDE TO AUTHORS

GENERAL

Please follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers for whom English is a second language and those who work in other fields. Please write clearly and simply, avoiding unnecessary technical terminology. *Nature's* staff will edit manuscripts to those ends if necessary. Contributors should check their proofs carefully. Because of the competition for space, many of the papers submitted for publication cannot be accepted. For this reason, and because brevity is a great assistance to readers, papers should be as brief as is consistent with intelligibility. Please note that one printed page of *Nature*, without diagrams or other interruptions of the text, has fewer than 1,300 words.

CATEGORIES OF PAPER

Review Articles survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance but must not exceed six pages of *Nature*.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*.

Articles should be accompanied by a heading of 50–80 words written to advertise their contents in general terms, to which editors will pay particular attention. A heading (printed in italic type) is not an abstract and should not usually contain numbers or measurements. The study should be introduced in more detail in the first two or three paragraphs, which should also briefly summarize its results and implications.

Articles may contain a few subheadings of two or three words. The meaning of the text should not depend on the subheadings, whose function is to break up the text and to point to what follows. There should be fewer than 50 references.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should not have more than 1,000 words of text or more than four display items and should not occupy more than two pages of *Nature*. The first paragraph should describe, in not more than 150 words, the origins and chief conclusions of the study. Letters should not have subheadings or more than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned, most are unsolicited. They are normally between one and four pages of *Nature* in length.

News and Views articles are intended to inform non-specialist readers about a recently published advance. Suggestions should be made to the News and Views Editor. Illustrations are welcome. Proposals for meeting reports should be agreed in advance.

Scientific Correspondence is for the discussion of scientific matters, including contributions published in *Nature*. Priority is given to contributions of less than 500 words and five references. Figures are welcome.

Submission. All manuscripts may be submitted to either London or Washington. They should not be addressed to editors by name. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT).

All contributions submitted for publication in *Nature* should conform with these rules.

Manuscripts should be typed, double-spaced, with a good-quality printer, on one side of the paper only. Four copies are required, each accompanied by lettered artwork. Four copies of half-tones should be included. Reference lists, figure legends and so on should be on separate sheets, all of which should be double-spaced and numbered. Copies of relevant manuscripts in press or submitted for publication elsewhere should be included, clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their reference numbers.

Titles should say what the paper is about with the minimum of technical terminology and in fewer than 80 characters. Authors should avoid active verbs, numerical values, abbreviations and punctuation and should include one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name, figure number and, when known, manuscript reference number. Original artwork should be unlettered. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are permitted only if the parts are closely related, either experimentally or logically. Suggestions for cover illustrations, with captions, are welcome. Original artwork (and one copy of the manuscript) will be returned when a manuscript cannot be published.

Colour Artwork. A charge of £500 a page is made for 4-colour figures. Inability to meet these costs will not prevent the publication of essential colour figures if the circumstances are explained.

Figure legends must not exceed 300 words and ideally should be much shorter, and should use telegraphic form wherever possible. The figure should be described first, then, briefly, the method. Reference to a method described elsewhere is preferable to a full description. Methods should not be described in detail in the text.

References should be numbered sequentially as they appear in the text, followed by those in tables and finally those in the figure legends. Reference numbers apply only to papers published or in the press, and a different number should be given to each paper cited. All other forms of reference (including unreferenced abstracts) should be included in the text as personal communication, manuscript in preparation or preprint (with number and institution where appropriate). Text should not be included in the reference list. References should be abbreviated according to the *World List of Scientific Periodicals* (fourth edition, Butterworth, London, 1963–65). The first and last page numbers should be cited. Reference to books should clearly indicate the publisher and the date and place of publication.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided as much as possible and, when used, defined. Editors will shorten words if necessary. In case of doubt, SI units should be used.

Footnotes should not be used except for changed addresses or to identify the corresponding author if different from the first named.

Acknowledgements should be brief. Grant numbers and contribution numbers are not allowed.

Using caged neurotransmitters

J. J. Marquet

The development of photolabile caged neurotransmitters has made possible the study of the split-second kinetics of receptor–ligand interactions. An instrument has now been developed for activating the neurotransmitter and measuring the neuron's response.

NICOTINIC acetylcholine receptor (AChR) is a protein that spans the plasma membrane of many nerve and muscle cell types. It plays a central role in the transmission of signals between these cells: acetylcholine that is released from a "firing" pre-synaptic cell diffuses across the synapse and binds to AChR molecules on the post-synaptic cell membrane. AChR responds to this binding process by changing its shape, on a sub-millisecond time-scale, to allow the rapid flow of sodium ions through the plasma membrane and into the post-synaptic cellular interior.

The AChR-acetylcholine complex changes its shape in a process named desensitization roughly one second later so that the initial high impedance of the plasma membrane to the flow of sodium ions is regained. The details of these shape changes are not well understood. *In vitro* studies of the kinetics of these processes have been constrained by the speed with which acetylcholine (and other neurotransmitters) can be applied to cell surfaces.

Over the past two decades, attempts have been made to circumvent these constraints using several techniques. Bartels and co-workers¹ introduced a photoisomerizable compound (3,3'-bis [(trimethylammonio) methylene] azobenzene bromide), the *cis* form of which at first appeared to be biologically inert and the *trans* form of which was found to be an active ligand for AChR. However, the *cis* form was later found to induce desensitization². An *in vitro* technique that uses a directed microscopic flow of acetylcholine analogues³ over cells successfully overcomes the problem of chemical artefact, but the kinetics of the resulting channel-opening current are convoluted with the kinetics of desensitization and the equilibration of agonist with AChR.

Cagey chemicals

The recent synthesis and characterization⁴ of a biologically inert photolabile precursor to carbamoylcholine (an AChR ligand) has renewed the possibility of using fast photochemistry to elucidate the fast opening kinetics of AChR. This new caged carbamoylcholine is one of several caged compounds (such as ATP⁵, inorganic phosphate⁶, and cyclic nucleotides⁷) that use a nitrophenyl protecting group covalently bound to the substrate of interest. Kinetic spectroscopic measure-

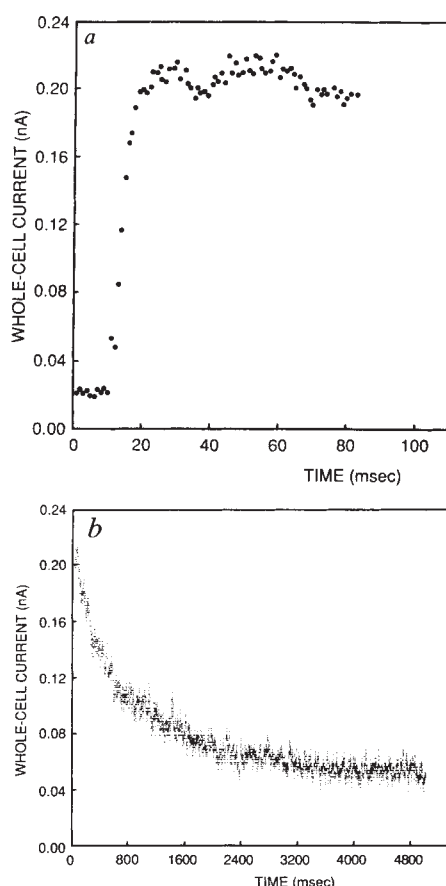


Fig. 1 Current influx into a transformed neuron versus time after laser photolysis of a caged carbamoylcholine. Both traces are from the same data set: only the time scales of display are different. Fluence at the neuron was 250 mJ cm⁻².

ments suggest that the recently reported caged carbamoylcholine photolyses, in the presence of ultraviolet light, to produce carbamoylcholine in less than 100 microseconds at physiological pH⁴.

The instrument

To exploit fully this caged neurotransmitter, an instrument has recently been developed⁸ that provides the capability to generate continuously tunable ultraviolet light, with a bandwidth of less than 1 nm, of sufficient fluence (energy per area), duration, and homogeneity to photolyse an arbitrary, but known, concentration of the caged carbamoylcholine. The instrument measures the average fluence of the ultraviolet light that irradiates the cell and the cell's surroundings, and can qualitatively evaluate the two-dimensional

fluence profile in the plane of the cell.

The instrument consists of a frequency-doubled flashlamp-pumped dye laser, an inverted fluorescence microscope, a patch-clamp apparatus⁹, and a computer fitted with a transient recorder. Steering optics guide most of the ultraviolet light from the laser, into the epi-port of the microscope, and through a quartz objective of the microscope; the light then strikes a cultured neuron that is bathed in caged neurotransmitter while voltage-clamped in the whole-cell configuration⁹. The remainder of the ultraviolet energy (about three per cent of the total) is diverted from the main beam and focused onto a pre-calibrated joule-meter from which the energy deposited at the cell neighborhood can be determined.

The area of the ultraviolet flash that strikes the cell neighborhood is determined by using a mirror whose metallized side has been scratched with a diamond pencil and placed on the microscope stage, facing the objective. After a scratch mark is brought into sharp focus in the field of view, an energetic (15 mJ) ultraviolet pulse is sent from the laser and through the microscope objective. The ultraviolet light vapourizes a microscopic circular area of the metal coating on the mirror, rendering a round spot transparent to the illuminator light of the microscope. The area of the spot is readily measured with a graticule, and the spot provides a permanent two-dimensional record of the sub-microsecond pulse of the laser in the focal plane of the objective.

The current traces shown in Fig. 1 were collected with the instrument described above. The data were collected from a transformed neuron that was patch-clamped in the whole-cell recording configuration and immersed in 100 μ M caged carbamoylcholine. The data were low-pass-filtered at 2.5 kHz before digitization. Figure 1a shows well-resolved channel-opening kinetics in the millisecond regime. Figure 1b shows the same data set displayed on a much longer time-scale and demonstrates the desensitization that is characteristic of AChR.

An issue that must be kept in mind when using this kind of technique is the potential problem of damage to the receptor and other cell constituents by ultraviolet radiation. It is essential to design control experiments that can detect the introduction of artefact into the

in vitro data. For example, the data that are shown above were collected only after it was demonstrated that the fluence used affected neither the amplitude of the response to agonist nor the overall impedance of the cell membrane⁸.

There are several powerful advantages associated with this new method. Both the bandwidth limit of the data acquisition system and the kinetics of the photolysis of the caged compound can be deconvoluted from the *in vitro* data, yielding the biological response to a temporal step-function of agonist. Furthermore, because the radiation fluence, the absorption cross-section for the caged agonist, and the quantum yield of photolysis⁴ are known, it is a simple matter to calculate the concentration of photo-released agonist present at the neuron. The optical configuration allows an extremely broad range of concentrations of caged agonist to be used with the same laser energies. These last two characteristics of the apparatus are important when studying the precise concentration dependence of the kinetics of receptor-mediated reactions over a wide range of ligand concentrations. □

Jeffrey Marque is at the Department of Engineering, Spinco Division, Beckman Instruments, 1050 Page Mill Road, Box 10200, Palo Alto, California 94303-0803, USA. For more information, fill in reader service number 100.

1. Bartels, E., Wassermann, N.H. & Erlanger, B.F. *Proc. natn. Acad. Sci. U.S.A.* **68**, 1820-1823 (1971).
2. Delcour, A.H. & Hess, G.P. *Biochemistry* **25**, 1793-1798 (1986).
3. Udgankar, J.B. & Hess, G.P. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8758-8762 (1987).
4. Milburn, T. *et al.* *Biochemistry* (in the press).
5. McCray, J.A., Herbette, L., Kihara, T. & Trentham, D.R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7237-7241 (1980).
6. Havinga, E., DeJongh, R.O. & Dorst, W. *Recl Trav. chim.* **75**, 378-383 (1956).
7. Engels, J. & Schlaeger, E.J. *J. med. Chem.* **20**, 907-911 (1977).
8. Marque, J. (in preparation).
9. *Single-channel Recording* (eds Sakmann, B. & Neher, E.) (Plenum, New York, 1983).

ADVERTISEMENT

Economical Modular Tissue Culture Incubator



- Safe for AIDS research • Reliable, for In-Vitro fertilization
- Versatile enough for thousands of lab uses

Easy to use — Just flush with gas mixture, seal and place at approp. temp. Perform multiple experiments simultaneously without cross contamination. Thousands in use worldwide.

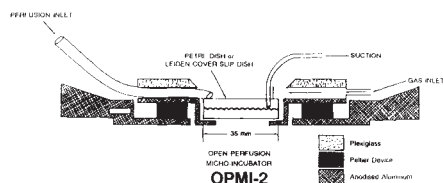
For information contact: **Flow Laboratories, Ltd.** or, **Billups-Rothenberg, Inc.** • P.O. Box 977
Del Mar, CA 92014-0977 USA • (619) 755-3309

Biophysics sampler

Several of the products to be displayed by the roughly 50 exhibitors at next week's American Biophysical Society meeting in Cincinnati, Ohio, are highlighted below.

Hi-TECH Scientific is out to make the technique of stopped-flow spectrophotometry accessible to every laboratory that has a spectrophotometer or a fluorometer. The company has announced that as of this month its universal **rapid kinetics accessory** will incorporate a new base material with enhanced chemical resistance (*Reader Service No. 101*). The modification makes its model SFA-11 accessory usable with all instruments that employ a 1-cm cuvette. Hi-Tech Scientific says its rapid kinetics accessory can be fitted in seconds and offers the ability to study reactions of less than 10 ms in duration, using only 200 µl of sample.

Medical Systems Corporation has an **open perfusion micro-incubator** which it says can produce culture conditions for cells in a perfused medium on the stage of any inverted microscope (*Reader Service No. 102*). The unit has built-in heating and cooling elements which control temperature by applying a d.c. current. It can maintain temperatures of between 8 °C and 46 °C under static conditions, or with flow rates of up to 2 ml min⁻¹. The company says a gas flow inlet across the



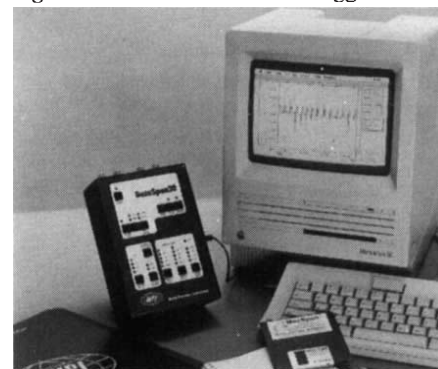
Medical Systems Corp. micro-incubator.

chamber allows for optimal temperature uniformity and pH control, and that optical and micro-mechanical access to the incubator are excellent. The micro-incubator can be used with disposable 35-mm petri dishes or with the company's Leiden oover slip dish, under high-magnification, oil immersion, phase and interference contrast, or fluorescence microscopy.

SLM Instruments is offering a new brochure on its Lifetime series of **spectrofluorometers** (*Reader Service No. 103*). The publication outlines the performance of both the SLM model 4800C scanning Lifetime spectrofluorometer and the SLM model 4800S multiple-frequency Lifetime spectrofluorometer, including use of the T-Optics design — which involves a light-path configuration with two emission paths symmetrically arranged on either side of the sample compartment — and

computer-assisted control and data acquisition. It includes optional configurations for the SLM 4800S and a performance comparison chart for both instruments. Two pages of the publication feature scans from a variety of research applications, and the brochure wraps up with a listing of the instruments' specifications and available accessories.

DataSpan20 is the name of a stand-alone **digital recorder and data logger** from



The DataSpan20 recorder and data logger.

World Precision Instruments which can interface with Macintosh and IBM PCs (*Reader Service No. 104*). WPI says data which has been digitized and stored in the RAM memory of DataSpan20 can be transferred to a personal computer by connecting the DataSpan20 to the computer's serial port, launching the computer application and selecting "Load". For data logging, the DataSpan20 can digitize and store analogue input at rates from one to 50,000 samples per second.

Narshige has recently launched a **precision microinjector** designed to deliver reproducible injections over periods of time (*Reader Service No. 105*). The model IM-200 has digital settings which allow the user to inject a specific volume of fluid — ranging from microlitres to femtolitres — repetitively through a pipette at variable intervals. The microinjector features adjustable fill and hold suctions, and injection, balance and clearing pressures. The injection can be triggered through a panel push-button, an optional foot switch, or a TTL logic pulse. A gating mode is available for external electrical or optional foot switch control timing. □

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further information, fill in the reader service card bound inside the journal.